



CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition,  
CC a method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 320 AA;

Query Match  
Best Local Similarity 100.0%; Score 1667; DB 7; Length 320;  
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPMTQVRYHGGAPLREVELPTCMYRLPMVHGRSYGAPAGAHVQERSNLSIQALESROD 60  
DB 1 MPMTQVRYHGGAPLREVELPTCMYRLPMVHGRSYGAPAGAHVQERSNLSIQALESROD 60  
QY 61 DIKRLYLELKAANDGSKMTIQTPDADLVNIIQADEPTTLTNALDLSVGYKGYALK 120  
DB 61 DIKRLYLELKAANDGSKMTIQTPDADLVNIIQADEPTTLTNALDLSVGYKGYALK 120  
QY 121 DIVYNNAPSPSLSLVHRLCEHFRVLSTVHTHSVSVBENLKCFGQNKQKOPROD 180  
DB 121 DIVYNNAPSPSLSLVHRLCEHFRVLSTVHTHSVSVBENLKCFGQNKQKOPROD 180  
QY 181 YOLGTTLIWKVPTQMKFSIQMCPLEGGINARFLFSLFGQKHNANATLIDSWDIA 240  
DB 181 YOLGTTLIWKVPTQMKFSIQMCPLEGGINARFLFSLFGQKHNANATLIDSWDIA 240  
QY 241 IFOLKESSSEKKAAPRSNMSAIGKSPMLAGNELTVADVVMSTVLOQIGGCVTVPANQ 300  
DB 241 IFOLKESSSEKKAAPRSNMSAIGKSPMLAGNELTVADVVMSTVLOQIGGCVTVPANQ 300  
QY 301 RMRSCENLAPFNTALKLK 320  
DB 301 RMRSCENLAPFNTALKLK 320

RESULT 2

ABM80843  
XX ABM80843 standard; protein, 320 AA.

AC ABM80843;  
XX 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) polypeptide PRO81501, SEQ:2179.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW cervical cancer; melanoma; leukemia; bladder cancer; pancreatic cancer;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.

OS Homo sapiens.  
XX  
XX WO2004030615-A2.  
XX  
XX 15-APR-2004.

XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PA (GETH) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
XX WPI; 2004-347921/32.  
DR N-PSDB; ACN38579.  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.

Claim 12; SEQ ID NO 2179; 7273bp; English.

XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acid and polypeptide  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention

Sequence 320 AA;

Query Match  
Best Local Similarity 100.0%; Score 1667; DB 8; Length 320;  
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPMTQVRYHGGAPLREVELPTCMYRLPMVHGRSYGAPAGAHVQERSNLSIQALESROD 60  
DB 1 MPMTQVRYHGGAPLREVELPTCMYRLPMVHGRSYGAPAGAHVQERSNLSIQALESROD 60  
QY 61 DIKRLYLELKAANDGSKMTIQTPDADLVNIIQADEPTTLTNALDLSVGYKGYALK 120  
DB 61 DIKRLYLELKAANDGSKMTIQTPDADLVNIIQADEPTTLTNALDLSVGYKGYALK 120  
QY 121 DIVYNNAPSPSLSLVHRLCEHFRVLSTVHTHSVSVBENLKCFGQNKQKOPROD 180  
DB 121 DIVYNNAPSPSLSLVHRLCEHFRVLSTVHTHSVSVBENLKCFGQNKQKOPROD 180  
QY 181 YOLGTTLIWKVPTQMKFSIQMCPLEGGINARFLFSLFGQKHNANATLIDSWDIA 240  
DB 181 YOLGTTLIWKVPTQMKFSIQMCPLEGGINARFLFSLFGQKHNANATLIDSWDIA 240  
QY 241 IFOLKESSSEKKAAPRSNMSAIGKSPMLAGNELTVADVVMSTVLOQIGGCVTVPANQ 300  
DB 241 IFOLKESSSEKKAAPRSNMSAIGKSPMLAGNELTVADVVMSTVLOQIGGCVTVPANQ 300  
QY 301 RMRSCENLAPFNTALKLK 320  
DB 301 RMRSCENLAPFNTALKLK 320

RESULT 3  
AAM25776

ID AAW25776 standard; protein; 312 AA.  
XX  
AC AAW25776;  
XX  
DT 19-DEC-1997 (first entry)  
XX  
DE JTV1 protein.  
XX  
DE JTV1 protein.  
XX  
KM JTV1; hPMS2; probe; detection; chromosome 7; deletion;  
KM mismatch repair gene; hereditary non-polyposis colorectal cancer;  
KM homologous recombination.  
XX  
OS Homo sapiens.  
XX  
XX MO9708312-A1.  
XX  
PD 06-MAR-1997.  
XX  
PF 26-AUG-1996; 96MO-US013598.  
XX  
PR 24-AUG-1995; 95US-00518862.  
XX  
PA (UYJO ) UNIV JOHNS HOPKINS.  
XX  
PI Vogelstein B, Kinzler KW, Nicolaides NC;  
XX  
DR WPI; 1997-179269/16.  
XX  
DR N-PSDB; AAT86182.  
XX  
PT Novel chromosome 7 gene, JTV1 - used for detecting chromosome 7  
XX  
PT deletions, and PMS2 promoter activity.  
XX  
PS Claim 5; Fig 2; 31pp; English.  
XX  
XX This sequence is JTV1 protein and is encoded by DNA isolated from human  
XX chromosome 7. The JTV1 coding sequence is located upstream from hPMS2.  
XX JTV1 cDNA can be used as probes to detect chromosome 7 deletions.  
XX involving JTV1. Due to the overlapping promoter regions, deletions of  
XX JTV1 would also affect PMS2 (a mismatch repair gene) expression, leading  
XX to hereditary non-polyposis colorectal cancer. JTV1 can also be used to  
XX assay activity or competence of the PMS2 promoter region, the presence of  
XX JTV1 suggesting that the PMS2 promoter is intact. JTV1 sequences can also  
XX be used to guide homologous recombination at the PMS2 locus  
XX  
SQ Sequence 312 AA;  
XX  
Query Match 97.8%; Score 1630; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEMYQVKPYHGGAPLRVLPCTCMYRLPNVHGSRSGYPAGAGHVQESNLSLQALESRD 60  
DB 1 MEMYQVKPYHGGAPLRVLPCTCMYRLPNVHGSRSGYPAGAGHVQESNLSLQALESRD 60  
QY 61 DILKRLYEKAAVDSKMIQTPDADLVNTIIQADEPTTLTTNALDLSVKGKYGALK 120  
DB 61 DILKRLYEKAAVDSKMIQTPDADLVNTIIQADEPTTLTTNALDLSVKGKYGALK 120  
QY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVHTHSSVKSVPENLLKCFGEQNKQPRD 180  
DB 121 DIVINANPASPPLSLVLRLLCEHFRVLSVHTHSSVKSVPENLLKCFGEQNKQPRD 180  
QY 181 YQUGFLIMKNVPTKTMKFSIQMCPISBEGNARFLFSIFGQKNAVNATLIDSVWDIA 240  
DB 181 YQUGFLIMKNVPTKTMKFSIQMCPISBEGNARFLFSIFGQKNAVNATLIDSVWDIA 240  
QY 241 IFOLKSGSSKEKAIVRSNMNSALGKSPMLAGNELTVADVVLMSVLIQOIGGCSVTVPANQ 300  
DB 241 IFOLKSGSSKEKAIVRSNMNSALGKSPMLAGNELTVADVVLMSVLIQOIGGCSVTVPANQ 300  
QY 301 RWRSCENTLAPF 312  
DB 301 RWRSCENTLAPF 312

RESULT 4  
ID ADR86551  
AC ADR86551 standard; protein; 312 AA.  
XX  
AC ADR86551;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE 1-312 amino acid sequence of p38/JTV-1 protein.  
XX  
DE p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.  
XX  
OS Homo sapiens.  
XX  
XX EP1454628-A2.  
XX  
PD 08-SEP-2004.  
XX  
PF 09-SEP-2003; 2003EP-00020344.  
XX  
PR 03-MAR-2003; 2003KR-00013058.  
XX  
PA (UYSE-) UNIV SEOUL NAT IND FOUND.  
XX  
PI Kim S, Park B;  
XX  
DR WPI; 2004-627822/61.  
XX  
DR N-PSDB; ADR86548.  
XX  
PT New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
XX e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
XX PT and cutaneous or intraocular melanoma, as well as for screening new  
XX anticancer agents.  
XX  
PS Claim 5; SEQ ID NO 4; 47pp; English.  
XX  
XX The present invention relates to an isolated p38/JTV-1 protein for use as  
XX medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
XX useful as medicament for treating breast cancer, large intestinal cancer,  
XX lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
XX cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
XX cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
XX rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
XX endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
XX Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
XX cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
XX cancer, urethral cancer, penile cancer, prostate cancer, chronic or acute  
XX leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
XX cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
XX CNS lymphoma, bone marrow tumour, brain stem nerve glioma, pituitary  
XX adenoma, or their combination. The protein is useful as a target for  
XX screening new anticancer agents. The present sequence represents the 1-  
XX 312 amino acid sequence of p38/JTV-1 protein.  
XX  
SQ Sequence 312 AA;  
XX  
Query Match 97.8%; Score 1630; DB 8; Length 312;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEMYQVKPYHGGAPLRVLPCTCMYRLPNVHGSRSGYPAGAGHVQESNLSLQALESRD 60  
DB 1 MEMYQVKPYHGGAPLRVLPCTCMYRLPNVHGSRSGYPAGAGHVQESNLSLQALESRD 60  
QY 61 DILKRLYEKAAVDSKMIQTPDADLVNTIIQADEPTTLTTNALDLSVKGKYGALK 120  
DB 61 DILKRLYEKAAVDSKMIQTPDADLVNTIIQADEPTTLTTNALDLSVKGKYGALK 120  
QY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVHTHSSVKSVPENLLKCFGEQNKQPRD 180  
DB 121 DIVINANPASPPLSLVLRLLCEHFRVLSVHTHSSVKSVPENLLKCFGEQNKQPRD 180





OS Homo sapiens.  
XX  
XX EPI454628-A2.  
XX  
XX 08-SEP-2004.  
XX  
XX 09-SEP-2003; 2003EP-00020344.  
XX  
XX 03-MAR-2003; 2003KR-00013058.  
XX  
XX (UYSE-) UNIV SEOUL NAT IND FOUND.  
XX  
XX Kim S, Park B;  
XX  
XX MPI; 2004-627822/61.  
XX N-PSDB; ADR86550.  
XX  
XX New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
PT e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
PT and cutaneous or intraocular melanoma, as well as for screening new  
PT anticancer agents.

PS Claim 5; SEQ ID NO 6; 47bp; English.

XX The present invention relates to an isolated p38/JTV-1 protein for use as  
XX medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
XX useful as medicament for treating breast cancer, large intestinal cancer,  
XX lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
XX cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
XX cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
XX rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
XX endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
XX Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
XX cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
XX tumor, urethral cancer, penile cancer, prostate cancer, chronic or acute  
XX leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
XX cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
XX CNS lymphoma, bone marrow tumour, brain stem nerve glioma, pituitary  
XX adenoma, or their combination. The protein is useful as a target for  
XX screening new anticancer agents. The present sequence represents the 84-  
XX 312 amino acid sequence of p38/JTV-1 protein.

XX Sequence 229 AA;

XX Query Match 71.4%; Score 1191; DB 8; Length 229;

XX Best Local Similarity 100.0%; Pred. No. 7, 2e-117; Indels 0; Gaps 0;

XX Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DADLVNTIIQADEPTTLTTNALDLSVSGDYGALKDVIINANPASPPLSLVLRLLC 143  
DB 1 DADLVNTIIQADEPTTLTTNALDLSVSGDYGALKDVIINANPASPPLSLVLRLLC 60  
QY 144 EHFRLVSTVHTSSVSVKVPENLLKCFGEONKKQPRDYOGLGFTLIWKVNPXTOMKESIQ 203  
DB 61 EHFRLVSTVHTSSVSVKVPENLLKCFGEONKKQPRDYOGLGFTLIWKVNPXTOMKESIQ 120  
QY 204 MCPPIGEGGNIARFLPSLFCQKKNANVATLIDSWVDIAIFOLKESGSKKCAVFRSNAL 263  
DB 121 MCPPIGEGGNIARFLPSLFCQKKNANVATLIDSWVDIAIFOLKESGSKKCAVFRSNAL 180  
QY 264 GKSPLMAGNELTVADVVLMSVLOQIGGCVTPANQVRMRSCENLAPF 312  
DB 181 GKSPLMAGNELTVADVVLMSVLOQIGGCVTPANQVRMRSCENLAPF 229

RESULT 7  
ADR86552  
ID ADR86552 standard; protein; 161 AA.

XX ADR86552;

XX 18-NOV-2004 (first entry)

XX

DE 1-161 amino acid sequence of p38/JTV-1 protein.

XX p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.

XX Homo sapiens.

XX EPI454628-A2.

XX 08-SEP-2004.

XX 09-SEP-2003; 2003EP-00020344.

XX 03-MAR-2003; 2003KR-00013058.

XX (UYSE-) UNIV SEOUL NAT IND FOUND.

XX Kim S, Park B;

XX MPI; 2004-627822/61.

XX N-PSDB; ADR86549.

XX New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
PT e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
PT and cutaneous or intraocular melanoma, as well as for screening new  
PT anticancer agents.

XX Claim 5; SEQ ID NO 5; 47bp; English.

XX The present invention relates to an isolated p38/JTV-1 protein for use as  
XX medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
XX useful as medicament for treating breast cancer, large intestinal cancer,  
XX lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
XX cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
XX cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
XX rectal cancer, anal cancer, colon cancer, vulval cancer, vaginal carcinoma,  
XX endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
XX Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
XX cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
XX tumor, urethral cancer, penile cancer, prostate cancer, chronic or acute  
XX leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
XX cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
XX CNS lymphoma, bone marrow tumour, brain stem nerve glioma, pituitary  
XX adenoma, or their combination. The protein is useful as a target for  
XX screening new anticancer agents. The present sequence represents the 1-  
XX 161 amino acid sequence of p38/JTV-1 protein.

XX Sequence 161 AA;

XX Query Match 49.9%; Score 832; DB 8; Length 161;

XX Best Local Similarity 100.0%; Pred. No. 3, 9e-79; Indels 0; Gaps 0;

XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMYQVKEPHYGGAFLRVELPTCMYRLPMVHGRSYGAPAGHVOESNLSIQALESRD 60  
DB 1 MEMYQVKEPHYGGAFLRVELPTCMYRLPMVHGRSYGAPAGHVOESNLSIQALESRD 60  
QY 61 DILKRLYEIKAAVNDGSKMTQTPDADLVNTIIQADEPTTLTTNALDLSVSGDYGALK 120  
DB 61 DILKRLYEIKAAVNDGSKMTQTPDADLVNTIIQADEPTTLTTNALDLSVSGDYGALK 120  
QY 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVHTSSVSV 161  
DB 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVHTSSVSV 161

RESULT 8  
ABP01502  
ID ABP01502 standard; protein; 51 AA.

XX ABP01502;

XX 24-JUN-2002 (first entry)

XX

DE Human ORFX protein sequence SEQ ID NO:2986.  
XX  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
XX MO200192523-A2.  
XX  
XX PD 06-DEC-2001.  
XX  
XX PF 29-MAY-2001; 2001WO-US010836.  
XX  
XX PR 30-MAY-2000; 2000US-0206132P.  
XX PR 29-AUG-2000; 2000US-0228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX PI Shinkets RA, Leach MD;  
XX WPI: 2002-106308/14.  
XX DR N-PSDB; ABN17254.  
XX  
XX PT Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and autoimmune disorders.  
XX  
XX PS Disclosure; SEQ ID NO 2986; 1037pp; English.  
XX  
XX CC The present invention describes substantially purified human proteins  
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
XX in the specification). ABN15762 to ABN27252 encode the human ORFX  
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
XX disorder or preventing a pathology associated with an ORFX-associated  
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
XX sequences can be used in gene therapy. ORFX sequences can be used in the  
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
XX osteoarthritis, neurodegenerative disorders, disorders related to organ  
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic  
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
XX storage disease, various immune deficiencies and disorders, infectious  
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also  
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,  
XX bone degenerative disorders, or periodontal disease, and for gut  
XX protection or regeneration and treatment of lung or liver fibrosis,  
XX retransfusion injury in various tissues and conditions resulting from  
XX systemic cytokine damage. N.B. The sequence data for this patent did not  
XX form part of the printed specification, but was obtained in electronic  
XX format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 51 AA;  
XX  
XX Query Match 14.8%; Score 247; DB 5; Length 51;  
XX Best Local Similarity 96.1%; Pred. No. 4.5e-18;  
XX Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 74 DGLSMQIOTPADIDVNNIIADDEPTLTNNALDLSVGLKGYGAKDII 124  
XX DB 1 DGLSMQIOTPADIDVNNIIADDEPTLTNNALDLSVGLKGYGAKDII 51  
XX  
XX RESULT 9  
XX ABB62468

ID ABB62468 standard; protein; 334 AA.  
XX  
XX AC ABB62468;  
XX  
XX XX 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 14196.  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX XX Drosophila melanogaster.  
XX  
XX XX MO200171042-A2.  
XX  
XX XX PD 27-SEP-2001.  
XX  
XX XX PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX  
XX PA (PEKE ) PE CORP NY.  
XX  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR WPI: 2001-656860/75.  
XX DR N-PSDB; ABL06571.  
XX  
XX XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.  
XX  
XX PS Disclosure; SEQ ID NO 14196; 21pp + Sequence Listing; English.  
XX  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB10140-AB116175) and the encoded proteins (AB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 334 AA;  
XX  
XX Query Match 12.9%; Score 214.5; DB 4; Length 334;  
XX Best Local Similarity 25.6%; Pred. No. 2.8e-13;  
XX Matches 85; Conservative 50; Mismatches 120; Indels 77; Gaps 10;  
XX  
XX QY 18 VELPTGMYRLPNV-----HRSYGPAPGAGHVOESNLSIQ----- 54  
XX DB 13 IKLPTCMYPLKXVSLAASLSGSSSTASTSCKLEANNRIDRGKAAATCALDLS 72  
XX QY 55 -----LESRODILKRYELKAAVDGSKMIGTPDDLDVNTIIOADEPT 99  
XX DB 73 LGRQIQRLKDPDTASVARKQVKLELKAQIQGIR-----AGLVCG--KTFQHT 123  
XX QY 100 TLTNNALDLSVGLKGYGAKDIIYNNAPSPSLIVLHRLCEHFRVLSVPHSSGVK 159  
XX DB 124 TAFQNG-----GLKEVPLQDVVINGHPFIPYALALKNAMRNLYTIDVTFTHSTWA 176  
XX QY 160 SV-----PENLLKCFGEONKQKOPRODYQLGFTLIMKRVPTQKFKSIQTMCPTEGSGN 212  
XX DB 177 DIGPAARFEANLAKV--PVNPALPK-----ISVTILMNCEHTEMISPTVYVYIGEVN 230  
XX QY 213 IARFLFSIFQCKNNAVATL---IDSWUDIAFQCKESGSKKAAVFRSNASALGKSPWL 269  
XX DB 231 IIRYLGIRVGAERYRSGSPLCNEIDLVDITCYQLLRCTMHTKTVAMWRLDLKRLQKQYF 290  
XX QY 270 AGNELTVADVYVLMVSVLQIQGCSVTVPANVOR 301

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:36:20 ; Search time 17.81 Seconds  
(without alignments)  
1341.256 Million cell updates/sec

Title: US-10-622-817-2

Perfect score: 1667  
Sequence: 1 MPWQVKRYHGGAGLRYEL.....RMRSCENLAPENTALKLK 320

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	100.0	341	4	US-09-949-016-11312
2	1630	97.8	312	2	US-08-518-862C-2
3	103	6.2	1512	3	US-09-443-184-48
4	97	5.8	559	2	US-08-756-317-10
5	97	5.8	559	4	US-09-091-609-4
6	97	5.8	5215	3	US-09-105-537-2
7	95	5.7	559	4	US-09-821-016-1
8	95	5.7	559	4	US-10-259-632-1
9	95	5.7	559	4	US-10-266-787-1
10	92	5.5	2954	4	US-09-150-867-1
11	91	5.5	474	4	US-09-248-796A-20321
12	90.5	5.4	359	3	US-09-540-824-2
13	90.5	5.4	597	1	US-08-399-696-102
14	88.5	5.3	1007	4	US-09-538-092-736
15	88	5.3	427	4	US-09-949-016-11178
16	88	5.3	724	4	US-09-900-920-62
17	88	5.3	883	4	US-09-976-232-2
18	88	5.3	883	4	US-10-289-779B-2
19	88	5.3	914	4	US-09-976-239-4
20	88	5.3	914	4	US-10-289-779B-4
21	86.5	5.2	761	3	US-09-012-710-13
22	86.5	5.2	761	3	US-09-556-273-13
23	86	5.2	211	4	US-09-328-352-6834
24	86	5.2	351	4	US-08-178-257-6
25	86	5.2	471	3	US-09-134-001C-4904
26	86	5.2	1365	3	US-09-376-330-18
27	86	5.2	1365	6	5194600-4

28	86	5.2	1365	6	5194600-4	Patent No. 5194600
29	85.5	5.1	1440	3	US-09-357-251-37	Sequence 37, Appl
30	85	5.1	559	4	US-10-253-509-1	Sequence 1, Appl
31	85	5.1	1135	2	US-08-469-537A-97	Sequence 97, Appl
32	84.5	5.1	579	4	US-09-107-433-4488	Sequence 4488, Ap
33	84.5	5.1	829	1	US-07-670-611-2	Sequence 2, Appl
34	84.5	5.1	829	1	US-08-220-674-2	Sequence 2, Appl
35	84.5	5.1	829	1	US-08-445-186-2	Sequence 2, Appl
36	84.5	5.1	829	1	US-08-446-548-2	Sequence 2, Appl
37	84.5	5.1	829	2	US-08-446-550-2	Sequence 2, Appl
38	83.5	5.0	778	4	US-09-583-110-3930	Sequence 3930, Ap
39	83.5	5.0	786	4	US-09-107-433-3893	Sequence 3893, Ap
40	83	5.0	559	2	US-09-052-339-1	Sequence 1, Appl
41	83	5.0	559	3	US-09-385-742B-1	Sequence 1, Appl
42	83	5.0	559	4	US-09-989-766-1	Sequence 1, Appl
43	83	5.0	767	3	US-08-836-567-8	Sequence 8, Appl
44	83	5.0	767	4	US-09-606-304-8	Sequence 8, Appl
45	82.5	4.9	222	4	US-09-248-796A-18146	Sequence 18146, A

## ALIGNMENTS

RESULT 1									
US-09-949-016-11312									
Sequence 11312, Application US/09949016									
Patent No. 6812339									
GENERAL INFORMATION:									
APPLICANT: VENTER, J. Craig et al.									
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED									
FILE REFERENCE: C1001107									
CURRENT APPLICATION NUMBER: US/09/949,016									
CURRENT FILING DATE: 2000-04-14									
PRIOR APPLICATION NUMBER: 60/241,755									
PRIOR FILING DATE: 2000-10-20									
PRIOR APPLICATION NUMBER: 60/237,768									
PRIOR FILING DATE: 2000-10-03									
PRIOR APPLICATION NUMBER: 60/231,498									
PRIOR FILING DATE: 2000-09-08									
NUMBER OF SEQ ID NOS: 207012									
SOFTWARE: FastSeq for Windows Version 4.0									
SEQ ID NO: 11312									
LENGTH: 341									
TYPE: PRT									
ORGANISM: Human									
US-09-949-016-11312									
Query Match									
Best Local Similarity 100.0%; Pred. No. 2.4e-181;									
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MEMYQVKRYHGGAGLRYELPTCMYRLPNVHGRSYGPAAGHVOESNLISQALESRD	60						
DB	22	MEMYQVKRYHGGAGLRYELPTCMYRLPNVHGRSYGPAAGHVOESNLISQALESRD	81						
QY	61	DILKLYELKAAVDSLSKICQTPDADLVNTIIOADEPTTLTNALDLSVSGKYGAK	120						
DB	82	DILKLYELKAAVDSLSKICQTPDADLVNTIIOADEPTTLTNALDLSVSGKYGAK	141						
QY	121	DIVINANPASPLSLIVLHRLICEHFRVLTSTHTSSSVKSVENLLKCGEQNKQPRD	180						
DB	142	DIVINANPASPLSLIVLHRLICEHFRVLTSTHTSSSVKSVENLLKCGEQNKQPRD	201						
QY	181	YOLGFTLIWKVVKPTQMKFSIQTMCPIGEGNIAFLFSLFCQKNAVNATLIDSWDIA	240						
DB	202	YOLGFTLIWKVVKPTQMKFSIQTMCPIGEGNIAFLFSLFCQKNAVNATLIDSWDIA	261						
QY	241	IFOLKESSKEKAAVFRSMNSALGKSPMLAGNELTVADVLMVLSVQIIGGCVTVPANQ	300						
DB	262	IFOLKESSKEKAAVFRSMNSALGKSPMLAGNELTVADVLMVLSVQIIGGCVTVPANQ	321						
QY	301	RMRSCENLAPENTALKLK 320							

Db 322 RMRSCEMLAPFNTALKLK 341

## RESULT 2

US-08-518-862C-2  
; Sequence 2, Application US/08518662C  
; Patent No. 5843757  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: Nicolaides, Nicholas C.  
; TITLE OF INVENTION: Human JTV1 Gene Overlaps PMS2 Gene  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/518, 862C  
; FILING DATE: 24-AUG-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 01107, 49697  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 312 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-518-862C-2

Query Match 97.8%; Score 1630; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 3.5e-177;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPMYQKPYHGGAPLRLVELPTCMYRLPNVHGRSYGAPAGAGHVOEESNLSIQLESROD 60  
Db 1 MPMYQKPYHGGAPLRLVELPTCMYRLPNVHGRSYGAPAGAGHVOEESNLSIQLESROD 60  
Qy 61 DILKRLYELKAAVNDGSKMIOPTPDADLDVNTIIOADEPTTLTNNALDINSVLGKDYALK 120  
Db 61 DILKRLYELKAAVNDGSKMIOPTPDADLDVNTIIOADEPTTLTNNALDINSVLGKDYALK 120  
Qy 121 DIVINANPASPPLSLVLRHLICEHFRVLSVTHSSVSKVSPENILKCFGQONKQKOPROD 180  
Db 121 DIVINANPASPPLSLVLRHLICEHFRVLSVTHSSVSKVSPENILKCFGQONKQKOPROD 180  
Qy 181 YOLGFTLLMKQVPTKQMFESIQTMCPLEGEGNINARFLPSLFGQKHNAVNATLIDSWDIA 240  
Db 181 YOLGFTLLMKQVPTKQMFESIQTMCPLEGEGNINARFLPSLFGQKHNAVNATLIDSWDIA 240  
Qy 241 IFOLKEGSSKKAAPFRSMNSALGKSPMLAGNELTVADVVLMSVLOOIGGCSVTVPANNO 300  
Db 241 IFOLKEGSSKKAAPFRSMNSALGKSPMLAGNELTVADVVLMSVLOOIGGCSVTVPANNO 300  
Qy 301 RMRSCEMLAPF 312  
Db 301 RMRSCEMLAPF 312

## RESULT 3

US-09-443-184-48  
; Sequence 48, Application US/09443184A  
; Patent No. 6372431  
; GENERAL INFORMATION:  
; APPLICANT: Cunningham, Mary Jane  
; APPLICANT: Zweiger, Gary  
; APPLICANT: Kaser, Matthew R.  
; APPLICANT: Panzer, Scott  
; APPLICANT: Selhammer, Jeffrey J.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Mariah  
; APPLICANT: Azimzai, Yalda  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS  
; FILE REFERENCE: PC-0007 US  
; CURRENT APPLICATION NUMBER: US/09/443, 184A  
; CURRENT FILING DATE: 1999-11-19  
; NUMBER OF SEQ ID NOS: 138  
; SOFTWARE: PERL Program  
; SEQ ID NO 48  
; LENGTH: 1512  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1  
US-09-443-184-48

Query Match 6.2%; Score 103; DB 3; Length 1512;  
Best Local Similarity 22.1%; Pred. No. 0.12; 69; Indels 60; Gaps 9;  
Matches 45; Conservative 30; Mismatches 69

Qy 124 INANPASPPLSLVLRHLICEHFRVLSVTHSSVSKVSPENILKCFGQONKQKOPRODYOL 183  
Db 6 LTVNSGDPPLGAL---LAVEHKDVS---SVEGKENILH----- 41  
Qy 184 GFTLLMKQVPTKQMFESIQTMCPLEGEGNINARFL-----PSLFGQKHNAVNATLIDSWV 237  
Db 42 ---VSENVIFTDV-----NSILKYLARVATTGLVGS--NLMHEHIDHWL 82  
Qy 238 DIAIFOLKEGSSKKAAPFRSMNSALGKSPMLAGNELTVADVVLMSVLOOIGGCS----- 292  
Db 83 EFSATKL--SSCDSFTSTINEINHLSTRTYLVGNSLSLADLCWATLK--GNAAMQEO 138  
Qy 293 ---VTVPANQRMRSCEMLAPFNT 314  
Db 139 KQKAPVHVKWFGLAQCAFOS 162

## RESULT 4

US-08-756-317-10  
; Sequence 10, Application US/08756317  
; Patent No. 5849894  
; GENERAL INFORMATION:  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Miskay, Timothy A.  
; APPLICANT: Stark, David M.  
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:57:26 ; Search time 50.2443 Seconds  
(without alignments)  
2084.158 Million cell updates/sec

Title: US-10-622-817-2

Perfect score: 1667  
Sequence: 1 MPMTGVKPYHGGAPLRYEL.....RWRBSCENTLAPFTALKLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBSCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBSCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBSCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBSCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBSCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBSCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBSCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBSCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBSCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBSCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBSCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10E\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBSCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1630	97.8	312	16 US-10-463-676-4	Sequence 4, Appli
2	1191	71.4	229	16 US-10-463-676-6	Sequence 5, Appli
3	832	49.9	161	16 US-10-463-676-5	Sequence 13717, A
4	102	6.1	201	15 US-10-369-493-13717	Sequence 92, Appli
5	101	6.1	826	10 US-09-746-660A-92	Sequence 90, Appli
6	101	6.1	833	10 US-09-746-660A-90	Sequence 5163, Ap
7	101	6.1	1221	9 US-09-919-891-2	Sequence 2, Appli
8	101	6.1	1221	9 US-09-738-626-5163	Sequence 2, Appli
9	101	6.1	1221	15 US-10-450-055-2	Sequence 156, App
10	100.5	6.0	1221	17 US-10-494-675-156	Sequence 58897, A
11	100.5	6.0	687	17 US-10-425-114-58897	Sequence 53993, A
12	100.5	6.0	1200	15 US-10-283-122A-53993	Sequence 3001, Ap
13	97.5	5.8	652	15 US-10-264-049-3001	

14	97.5	5.8	711	16 US-10-437-963-175483	Sequence 175483, A
15	97	5.8	559	9 US-09-364-847-21	Sequence 21, Appli
16	97	5.8	856	9 US-09-364-847-13	Sequence 33, Appli
17	97	5.8	856	9 US-09-364-847-13	Sequence 35, Appli
18	97	5.8	5215	9 US-09-861-289-2	Sequence 2, Appli
19	97	5.8	5215	9 US-09-860-846-2	Sequence 2, Appli
20	97	5.8	5215	10 US-09-988-348A-2	Sequence 2, Appli
21	97	5.8	5215	10 US-09-836-821-2	Sequence 45, Appli
22	97	5.8	5215	14 US-10-271-889-45	Sequence 153870, A
23	96.5	5.8	914	16 US-10-437-963-153870	Sequence 29, Appli
24	96.5	5.8	1148	16 US-10-655-799-29	Sequence 115189, A
25	96	5.8	606	16 US-10-437-963-115189	Sequence 42533, A
26	95.5	5.7	455	15 US-10-425-114-42633	Sequence 1, Appli
27	95	5.7	559	9 US-09-821-016-1	Sequence 1, Appli
28	95	5.7	559	9 US-09-820-952A-1	Sequence 1, Appli
29	95	5.7	559	14 US-10-218-519-1	Sequence 1, Appli
30	95	5.7	559	14 US-10-259-632-1	Sequence 1, Appli
31	95	5.7	559	14 US-10-266-787-1	Sequence 1, Appli
32	95	5.7	559	14 US-10-252-518-1	Sequence 1, Appli
33	95	5.7	559	17 US-10-914-244-1	Sequence 1, Appli
34	95	5.7	722	15 US-10-369-493-10013	Sequence 10013, A
35	95	5.7	936	14 US-10-032-585-7442	Sequence 7442, Ap
36	94.5	5.7	955	15 US-10-282-122A-69912	Sequence 69912, A
37	94	5.6	203	15 US-10-369-493-8604	Sequence 8604, Ap
38	94	5.6	468	15 US-10-424-599-260388	Sequence 260388, A
39	94	5.6	1440	16 US-10-437-963-136277	Sequence 136277, A
40	93.5	5.6	569	16 US-10-437-963-123904	Sequence 123904, A
41	93.5	5.6	733	15 US-10-424-599-243766	Sequence 243766, A
42	92	5.5	739	15 US-10-424-599-251522	Sequence 251522, A
43	92	5.5	1206	15 US-10-283-122A-63838	Sequence 63838, A
44	91.5	5.5	1149	16 US-10-655-799-10	Sequence 30, Appli
45	91.5	5.5	1192	15 US-10-283-122A-64644	Sequence 64644, A

#### ALIGNMENTS

RESULT 1  
US-10-463-676-4  
; Sequence 4, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; TITLE OF INVENTION: for Screening Pharmaceutical Composition for Treating Cancer  
; FILE REFERENCE: 01679-09  
; CURRENT APPLICATION NUMBER: US/10/463, 676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; PRIOR FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(312)  
; OTHER INFORMATION: 1-312 amino acid sequence of p38/JTV-1  
US-10-463-676-4  
Query Match 97.8%; Score 1630, DB 16; Length 312;  
Best Local Similarity 100.0%; Pred. No. 6.4e-153;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPMTGVKPYHGGAPLRYELPTCMRLPYHRSYGPAGAGHVEESLSTQALESSRD 60  
DB 1 MPMTGVKPYHGGAPLRYELPTCMRLPYHRSYGPAGAGHVEESLSTQALESSRD 60  
QY 61 DIKLRYELKAVDGLSKMIGTPDADLVNTIIOADEPTLTNNALDLSVGLKGYGALK 120

Db 61 DILKRLYLKAAVDSKMLIQFPDADLDVTNIIQADEPTTLTNALDINSVLGKDYALK 120  
Qy 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVTHSSVSVSPENLLKCGEONKKQPRD 180  
Db 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVTHSSVSVSPENLLKCGEONKKQPRD 180  
Qy 181 YOLGFTLWKVNPYKTMKFSIQTMCPIEBGNARFLFSFGQKNAVANATLIDSWVDA 240  
Db 181 YOLGFTLWKVNPYKTMKFSIQTMCPIEBGNARFLFSFGQKNAVANATLIDSWVDA 240  
Qy 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVLVMSVLOQIGCSVTVPANVQ 300  
Db 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVLVMSVLOQIGCSVTVPANVQ 300  
Qy 301 RMRSCENLAPF 312  
Db 301 RMRSCENLAPF 312

## RESULT 2

US-10-463-676-6  
; Sequence 6, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; APPLICANT: Park, Bum-Joon  
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FaestSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(229)  
; OTHER INFORMATION: 84-312 amino acid sequence p38/JTV-1  
US-10-463-676-6

Query Match 71.4%; Score 1191; DB 16; Length 229;  
Best Local Similarity 100.0%; Pred. No. 1.5e-109;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 DADLDVTNIIQADEPTTLTNALDINSVLGKDYALKDIVINANPASPPLSLVLRLLC 143  
Db 1 DADLDVTNIIQADEPTTLTNALDINSVLGKDYALKDIVINANPASPPLSLVLRLLC 60  
Qy 144 EHFRLVSTVHTSSVSVSPENLLKCGEONKKQPRDYOGLTLLMKVNPYKTMKFSIQ 203  
Db 61 EHFRLVSTVHTSSVSVSPENLLKCGEONKKQPRDYOGLTLLMKVNPYKTMKFSIQ 120  
Qy 204 MCPITGEGNINARFLFSFGQKNAVANATLIDSWVDAIFOLKEGSSKEKAAVFRSMNSAL 263  
Db 121 MCPITGEGNINARFLFSFGQKNAVANATLIDSWVDAIFOLKEGSSKEKAAVFRSMNSAL 180  
Qy 264 GKSPLMAGNELTVADVLVMSVLOQIGCSVTVPANVQMRMRSCEMLAPF 312  
Db 181 GKSPLMAGNELTVADVLVMSVLOQIGCSVTVPANVQMRMRSCEMLAPF 229

## RESULT 3

US-10-463-676-5  
; Sequence 5, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; APPLICANT: Park, Bum-Joon

; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FaestSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 161  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(161)  
; OTHER INFORMATION: 1-161 amino acid sequence of p38/JTV-1  
US-10-463-676-5

Query Match 49.9%; Score 832; DB 15; Length 161;  
Best Local Similarity 100.0%; Pred. No. 3.9e-74;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPMYVKEPYHGGAPLRYELPTCMYRLPNVHGRSGYPAGAGHYOESNLSLOALESRD 60  
Db 1 MPMYVKEPYHGGAPLRYELPTCMYRLPNVHGRSGYPAGAGHYOESNLSLOALESRD 60  
Qy 61 DILKRLYLKAAVDSKMLIQFPDADLDVTNIIQADEPTTLTNALDINSVLGKDYALK 120  
Db 61 DILKRLYLKAAVDSKMLIQFPDADLDVTNIIQADEPTTLTNALDINSVLGKDYALK 120  
Qy 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVTHSSVSVK 161  
Db 121 DIVINANPASPPLSLVLRLLCEHFRVLTSTVTHSSVSVK 161

## RESULT 4

US-10-369-493-13717  
; Sequence 13717, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xiandeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 2002-02-21  
; SEQ ID NO 13717  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Pseudomonas fluorescens  
US-10-369-493-13717

Query Match 6.1%; Score 102; DB 15; Length 201;  
Best Local Similarity 30.5%; Pred. No. 0.17;  
Matches 32; Conservative 15; Mismatches 36; Indels 22; Gaps 3;

Qy 211 GNIRFLPSLFGQKNAVNA-TLIDSWVDAIFOLKEGSSKEKAAVFRSMNSALGKSPWL 269  
Db 106 GPAARLITVGAFAENAEVITRAHNFVKV-----MDLEIGKTPYL 146  
Qy 270 AGNELTVADVLVMSVLOQI--GGSVTVPANVQMRMRSCEMLAPF 312  
Db 147 AGTEPTIADVSAYSIVIAHAPGNVSLDDYANVRAMLRIEALPGF 191

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:34:50 / Search time 13.3213 seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-2

Perfect score: 1667

Sequence: 1 MEMYGVKPYHGAGAPLRLVEL.....RWMRSCEMLAPFTALKLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :  
1: pir\_79: \*  
2: pir1: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	7.5	719	2	T52043
2	111	6.7	728	2	T01200
3	105.5	6.3	222	2	G82441
4	97.5	5.8	218	2	A29036
5	97.5	5.8	645	2	JC5517
6	97	5.8	559	2	A38604
7	96.5	5.8	174	2	S74241
8	96	5.8	1675	2	S70770
9	94	5.6	218	2	B29231
10	93	5.6	393	2	D64391
11	92	5.5	210	2	B85594
12	92	5.5	210	2	F64821
13	92	5.5	210	2	P90743
14	92	5.5	744	2	AE2703
15	92	5.5	764	2	E97485
16	92	5.5	867	2	S72842
17	92	5.5	1036	2	AF2338
18	92	5.5	1206	2	E87072
19	92	5.5	2954	2	T14156
20	91.5	5.5	218	1	XURG4
21	91.5	5.5	1192	2	G70513
22	91	5.5	2279	2	T42531
23	91	5.5	2280	2	T38906
24	90.5	5.4	359	2	T37921
25	90	5.4	710	2	S67098
26	89.5	5.4	218	2	B34159
27	89.5	5.4	1365	1	BVBK5
28	89	5.3	2480	2	D84904
29	88.5	5.3	209	1	XUPE11

30	88.5	5.3	1007	2	S48535	rho-type GTPase-ac
31	88	5.3	207	2	D72272	conserved hypochet
32	88	5.3	218	2	B28946	glutathione transf
33	88	5.3	437	2	I51238	translacion elonga
34	87.5	5.2	214	2	B46681	glutathione transf
35	87.5	5.2	1065	2	A43421	SEC8 protein - yea
36	87.5	5.2	1292	2	T31462	probable magnesium
37	87.5	5.2	1714	1	S16644	multifunctional am
38	87.5	5.2	2586	2	T21676	hypothetical prote
39	87	5.2	586	2	C71695	ctp synthase (pyrG
40	86.5	5.2	208	1	S43851	glutathione transf
41	86.5	5.2	379	2	D97198	methyl-accepting c
42	86.5	5.2	436	2	I51237	translacion elonga
43	86.5	5.2	439	2	D86440	unknown protein [I
44	86.5	5.2	749	2	A57276	MAD1 protein - yea
45	86	5.2	377	2	F69172	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T52043  
probable glutamate-tRNA ligase (EC 6.1.1.17) (imported) - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #ext\_change 09-Jul-2004  
C/Accession: T52043

R/day, I.S.; Golovkin, M.; Reddy, A.S.  
Biochim. Acta 1399, 219-24, 1998

A/Title: Cloning of the cDNA for glutamyl-tRNA synthetase from Arabidopsis thaliana.

A/Reference number: Z24836; MUID:9765600; PMID:9765600

A/Accession: T52043

A/status: preliminary; translated from GB/EMBL/DBJ

A/molecule type: mRNA

A/Residues: 1-719 <DAY>

A/Cross-references: UNIPROT:O82462; EMBL:AF067773; PIDD: AAC36469.1

A/Suprafamily: Yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology

C/Keywords: ligase

Query Match 7.5%; Score 125.5; DB 2; Length 719;  
Best Local Similarity 27.3%; Pred. No. 0.017;

Matches 48; Conservative 26; Mismatches 53; Indels 49; Gaps 9;

QY 128 PASPLSLVLRILCEHVRVSLVTHSSVSVENLKCCEGQKQPRDYGFTL 187

DB 10 PESPPLSVVAUSLSASPV---TIDSSNAATVPSFV---PSDGKLN-----GATV 55

QY 188 IKNVPEKTOMKPSIQTMCPIDEGGNIAFLPSLFGQKHNAVATLIDSWVDIA-IFOLKE 246

DB 56 LARVY-----GNSAKKLPPFYG--NNAPSSQIDENVDAVSF--SS 93

QY 247 GSSKEKAIVFRSMNSALGKSPWLAGNELTVADVLTMSVLOQIGGCVTPPANQVR 302

DB 94 GSEFENAC--GRVDKYLESTPLVGHSHSIDVAIVASLAGTG-----QRW 137

##### RESULT 2

T01200  
probable glutamate-tRNA ligase (EC 6.1.1.17) F21E10.12 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #ext\_change 09-Jul-2004  
C/Accession: T01200

R/Davieson, S.; Rohlfing, T.; David, M.; O'Brian, D.

submitted to the EMBL Data Library, April 1998.

A/Description: The sequence of A. thaliana F21E10.

A/Reference number: Z14258

A/Accession: T01200

A/status: translated from GB/EMBL/DBJ

A/molecule type: DNA

A/Residues: 1-728 <DAY>

A/Cross-references: UNIPROT:O65253; EMBL:AF058914; NID:G3047074; PIDD:G3047084; GSPDB:GNC

A/Experimental source: cultivar Columbia

C/Genetics:

A:Gene: ATSP:F21E10.12  
A:Map position: 5  
A:Introns: 47/2; 89/3; 141/1; 503/3; 659/3  
C:Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
F:223-499/Domain: glutamine-tRNA ligase homology <EGL>

Query Match 6.7%; Score 111; DB 2; Length 728;  
Best Local Similarity 25.9%; Pred. No. 0.3;  
Matches 48; Conservative 26; Mismatches 53; Indels 58; Gaps 10;

QY 128 PASPPLSLVLRLLCEHFRLVSTVHTSVKSVPEMLKCGEQNKQPRODYOLGFTL 187  
DB 10 PESPLSVYVALSLISASPV-----TIDSSAATVPSPFV---FSDGRKLN-----GATV 55  
QY 188 IMKNVPTQMKFSIQTMPIEGEINARFLSLFGQKNNVNAI-----LIDSWD 238  
DB 56 LIRTV-----GRSAKLPDPFG--NNAFDSQVAILCINMKIDEMVD 95  
QY 239 IAFOLKESGSKKAAVFRSMNSALGKSPWLAGNELTVADVLMASVLAQIGGCSVTVPA 297  
DB 96 YASVF--SSGSEFENAC--GRVDKYLSSSTFLVGHSLSIADVALWSALAGTG----- 143  
QY 298 NVQRM 302  
DB 144 --QRM 146

RESULT 3  
G82441  
probable glutathione S-transferase VCA0584 [imported] - Vibrio cholerae (strain N16961)  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C/Accession: G82441  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gelin, M.L.; Dodson, R.J.;  
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: G82441  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-222 <HET>  
A:Cross-references: UNIPROT:Q9RM05; GB:AB004389; GB:AE003853; NID:g96657989; PIDN:AAF9648  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Gene: VCA0584  
A:Map position: 2  
C:Superfamily: hypothetical protein b2302

Query Match 6.3%; Score 105.5; DB 2; Length 222;  
Best Local Similarity 25.7%; Pred. No. 0.17;  
Matches 28; Conservative 22; Mismatches 38; Indels 21; Gaps 4;

QY 197 MKFSIQTMPIEGEINARFLSLFGQKNNVNAIIDLWDVLAIFOLKESGSKKAAY 256  
DB 97 LMFQMGVGPMMGQANV---FTRYPEKIQPA---IDRY-----QKGRRLF 137  
QY 257 RSMNSALGKSPWLAGNELTVADVLM--WSVLQOIGGCSVTVPANVQRM 303  
DB 138 EVMDDQLAQNPLYAGDEYTIADATFPVVRHHEMGSISIDGLTHLQRMK 186

RESULT 4  
A29036  
glutathione transferase (EC 2.5.1.18) Yb3 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 09-Jul-2004  
C/Accession: A29036  
R:Abramowitz, M.; Listowsky, I.  
J. Biol. Chem. 262, 7770-7773, 1987  
A:Title: Selective expression of a unique glutathione S-transferase Yb3 gene in rat brain

A:Reference number: A29036; MUID:87222405; PMID:3584141  
A/Accession: A29036  
A:Molecule type: mRNA  
A:Residues: 1-218 <ABR>  
A:Cross-references: UNIPROT:P08009; GB:J02744; NID:g204512; PIDN:AAA41292.1; PID:g20451  
C:Superfamily: glutathione transferase  
C:Keywords: transferase

Query Match 5.8%; Score 97.5; DB 2; Length 218;  
Best Local Similarity 25.6%; Pred. No. 0.78;  
Matches 33; Conservative 24; Mismatches 41; Indels 31; Gaps 5;

QY 179 QYQVGFLLIMKNVPTQMKFSIQTMCPREGGINARFLSLFGQKNNVNAI----- 231  
DB 49 EKFKLG--LDFPNLP-----YLIDSHKITQSNALIRYL---GRKNHLCGTEBERIRV 97  
QY 232 -----LIDSWDIA-----IFOLKESGSKKAAVFRSMNSALGKSPWLAGNELTVAD 278  
DB 98 DILENQMDNRMVAVLARLCNPDPEKLRQYELQPLGMKRLYSEFLGKRPWAGDKITVD 157  
QY 279 VFLMSVLAQ 287  
DB 158 FIAYDVLER 166

RESULT 5  
JC5517  
Gu/RNA helicase II binding protein - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 05-Nov-1999  
C/Accession: JC5517  
R:Valdez, B.C.; Henning, D.; Perlaky, L.; Busch, R.K.; Busch, H.  
Biochem. Biophys. Res. Commun. 234, 335-340, 1997  
A:Title: Cloning and characterization of Gu/RH-II binding protein.  
A:Reference number: JC5517; MUID:97320420; PMID:9177271  
A/Accession: JC5517  
A:Molecule type: mRNA  
A:Residues: 1-645 <VAL>  
A:Cross-references: GB:U78524; NID:g1696006; PIDN:AA58488.1; PID:g1696007  
A:Note: It is uncertain whether Met-4 or Met-6 is the initiator  
C:Comment: This protein is localized to the nucleus and interacts with Gu/RNA helicase I  
F:50-58/Region: nuclear location signal  
F:362-374/Region: nuclear location signal  
F:514-517,551-554,592-595,606-609/Region: 4-residue repeats (N-T-S-L)  
F:7,177,322,412,460,461,462,467,614,627/Binding site: phosphate (Ser) (covalent) #sc  
F:65,394,624/Binding site: phosphate (Thr) (covalent) #scatue predicted

Query Match 5.8%; Score 97.5; DB 2; Length 645;  
Best Local Similarity 20.5%; Pred. No. 3.7;  
Matches 79; Conservative 56; Mismatches 129; Indels 121; Gaps 20;

QY 16 LRV-ELPTCM-YRLPNVHGRSYGPARGAGHVOESNLSLOALSRDDILKRLYEKAAV 73  
DB 8 LRVSELQVLLGYAGRNHGRKHELLTYALHLRAG--CSPAVQMKIKELRYRRFPQK--- 62  
QY 74 DGLSKMIQTPDADLDVTNIIQADEPTLLITNALDNLVSGDKYGALIKDIYINNPASPL 133  
DB 63 -----IMTP-ADLSIPNVHSSPMPTLSPS-----TIPQLYTDGHPASSPL 102  
QY 134 ---SL-VLRLLCEHF-RVLSTVHTSSVSVY-----ENLIK-CFGEQNKQ----- 176  
DB 103 LPVSLGPKHELELPHTLSALHPVHPDILQKLPFYDLIDELIKPTSLASDNGRFRRTG 162  
QY 177 -----PROYQVGFLLIMKNVPTQMKFSIQ-----TMCPTEGG----- 211  
DB 163 FARALTPQVQVQVQSSSM--DISGTCDFTVQVQLRFLCSETSCPEQDHPPPNLCKVNT 219  
QY 212 -----NIARFLSLFGQKNNVNAIIDLSDV-----DI 239  
DB 220 KPCSLPGYLPPTKNGVEPKRSPRINITSLV-----RLSTTVPTNTIVSWTAEIGNYGM 274  
QY 240 AIRQLKESG-----KKAIVFRSMNSALGKSPWLAGNELTVADVLM--WSVLQOIGGC 291



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:33:04 / Search time 60.8145 Seconds  
(without alignments)  
2694.512 Million cell updates/sec

Title: US-10-622-817-2

Perfect score: 1667

Sequence: 1 MEMYGVKPYHGGAPLRLVEL.....RWRSCENLAPFNTALKLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667	100.0	320	1	MCA2_HUMAN
2	1663	99.8	320	2	Q96CZ5
3	1464	87.8	320	2	Q8R010
4	1464	87.8	320	2	Q8R2Y6
5	1444	86.6	320	1	MCA2_CRIGR
6	1260	75.6	280	2	Q8R3V2
7	1144.5	68.7	311	2	Q6DK86
8	1139.5	68.4	311	2	Q6INU4
9	1132.5	67.9	311	2	Q7ZYD7
10	887.5	53.2	321	2	Q7Y3C0
11	267.5	16.0	340	2	Q7Q7A3
12	224	13.4	301	2	Q7KUM5
13	224	13.4	313	2	Q8T060
14	224	13.4	322	2	Q6NKM4
15	214.5	12.9	334	1	MCA2_DROME
16	125.5	7.5	719	2	Q82462
17	111	6.7	728	2	Q65253
18	106.5	6.4	880	2	Q8ERT1
19	106.5	6.4	913	2	Q6DRB3
20	106.5	6.4	913	2	Q6PRQ4
21	105.5	6.3	222	2	Q9KM05
22	105	6.3	240	2	Q8BTR1
23	105	6.3	1512	1	SYEP_MOUSE
24	105	6.3	1800	2	Q7PRA2
25	104.5	6.3	922	2	Q6TMS5
26	104	6.2	661	2	Q9VB85
27	104	6.2	702	2	Q8IRC6
28	104	6.2	328	2	Q8IMQ2
29	103	6.2	328	2	Q86X73
30	103	6.2	869	2	Q6PD57
31	103	6.2	1213	2	Q8FTD2

32	102.5	6.1	372	2	Q87XN7	Q87XN7 pseudomonas
33	102.5	6.1	715	2	Q9LIZ8	Q9LIZ8 oryza sativ
34	102	6.1	559	2	Q8KQ23	Q8KQ23 pseudomonas
35	102	6.1	559	2	Q8BD25	Q8BD25 pseudomonas
36	102	6.1	1486	2	Q6TXE9	Q6TXE9 ratius norv
37	101	6.1	1196	2	Q6M580	Q6M580 corynebacte
38	101	6.1	1221	2	Q8NQD1	Q8NQD1 corynebacte
39	100.5	6.0	454	2	Q6UDR4	Q6UDR4 mus musculu
40	100.5	6.0	1201	2	Q6NH83	Q6NH83 corynebacte
41	99.5	6.0	651	1	PIA1_MOUSE	Q89307 mus musculu
42	99	5.9	203	2	Q6W1I2	Q6W1I2 rhizobium s
43	98.5	5.9	222	2	Q8EGM4	Q8EGM4 shewanella
44	98.5	5.9	582	2	Q6ECB1	Q6ECB1 yersinia ps
45	97.5	5.8	217	1	GTW3_RAT	P08009 rattus norv

## ALIGNMENTS

RESULT 1  
MCA2\_HUMAN STANDARD; PRT; 320 AA.  
ID Q13155, Q9P1L2;  
AC Q13155, Q9P1L2;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Multisynthetase complex auxiliary component p38 (JTV-1 protein)  
DE (PRO0992).  
GN Name=JTV1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96115582; PubMed=8666379;  
RA Nicolaidis N.C., Kinzler K.W., Vogelstein B.;  
RT "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts  
and a novel overlapping gene";  
RL Genomics 29:329-334 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshuler S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP SEQUENCE OF 197-320 FROM N.A.  
RC TISSUE=Fetal liver;  
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
RA Liu M., He P.;  
RT "Functional prediction of the coding sequences of 121 new genes  
deduced by analysis of cDNA clones from human fetal liver";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]

RP INTERACTION WITH FUBP1.  
RX MEDLINE=22716800; PubMed=12819782; DOI=10.1038/ng1182;  
RA Kim M.J., Park B.-J., Kang Y.-S., Kim H.J., Park J.-H., Kang J.W.,  
RT Lee S.W., Han J.W., Lee H.-W., Kim S.,  
RT "Downregulation of FUSE-binding protein and c-myc by tRNA synthetase  
RT cofactor p38 is required for lung cell differentiation.";  
RL Nat. Genet. 34:330-336(2003).  
CC -1- FUNCTION: Probable core protein of the multisynthetase complex  
CC structure. Mediates ubiquitination of FUBP1 and its degradation by  
CC the proteasome.  
CC -1- SUBUNIT: Component of the multisynthetase complex which is  
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the  
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary  
CC proteins, p18, p48 and p43. Binds FUBP1.  
CC -1- SIMILARITY: Contains 1 GST-like domain.  
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 312.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, U24169; AAC50391.1; ALT FRAME.  
DR EMBL, BC002853; AAH02853.1; -  
DR EMBL, BC010156; AAH01056.1; -  
DR EMBL, AF116615; AAF71039.1; -  
DR H-InvDB; HIX0006460; -  
DR MIM; 600859; -  
DR InterPro: IPR010987; GST\_C like.  
DR InterPro: IPR004046; GST\_Cterm.  
DR Pfam: PF00043; GST\_C\_1.  
DR KX Protein biosynthesis.  
SQ SEQUENCE 320 AA; 35349 MW; F253726B63C12BAB CRC64;  
Query Match 100.0%; Score 1667; DB 1; Length 320;  
Best Local Similarity 100.0%; Pred. No. 6e-127;  
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPWYQVRYHGGAGPLRYELPTCMYRLPNVHGSRGYPAPGAGHVOESNLSLOALESROD 60  
DB 1 MPWYQVRYHGGAGPLRYELPTCMYRLPNVHGSRGYPAPGAGHVOESNLSLOALESROD 60  
QY 61 DILRLYEIKAAVDSLSKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
DB 61 DILRLYEIKAAVDSLSKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
DB 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
QY 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
DB 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
QY 181 YOLGFTLLIMKNVPTKQMKFSIQTMCPTEGEGNIARFLFSLFGQKNAVNAATLIDSWDIA 240  
DB 181 YOLGFTLLIMKNVPTKQMKFSIQTMCPTEGEGNIARFLFSLFGQKNAVNAATLIDSWDIA 240  
QY 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVILMSVLOQIGGCSVTVPANVQ 300  
DB 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVILMSVLOQIGGCSVTVPANVQ 300  
QY 301 RWRMSCENLAPFNLTALKLK 320  
DB 301 RWRMSCENLAPFNLTALKLK 320  
RESULT 2  
Q96CZ5 PRELIMINARY; PRT; 320 AA.  
AC Q96CZ5;

DT 01-DEC-2001 (TRENBLREL 19, Created)  
DT 01-DEC-2001 (TRENBLREL 19, Last sequence update)  
DT 01-MAR-2004 (TRENBLREL 26, Last annotation update)  
DE JTV1.  
GN Name=JTV1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.U., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Locantello N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz U., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Maitra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Straubeberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, BC013630; AAH13630.1; -  
DR InterPro: IPR004046; GST\_Cterm.  
DR InterPro: IPR010987; GST\_C like.  
DR Pfam: PF00043; GST\_C\_1.  
SQ SEQUENCE 320 AA; 35335 MW; 19F14BF758612E08 CRC64;  
Query Match 99.8%; Score 1663; DB 2; Length 320;  
Best Local Similarity 99.7%; Pred. No. 1.3e-126;  
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MPWYQVRYHGGAGPLRYELPTCMYRLPNVHGSRGYPAPGAGHVOESNLSLOALESROD 60  
DB 1 MPWYQVRYHGGAGPLRYELPTCMYRLPNVHGSRGYPAPGAGHVOESNLSLOALESROD 60  
QY 61 DILRLYEIKAAVDSLSKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
DB 61 DILRLYEIKAAVDSLSKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
DB 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
QY 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
DB 121 DIVINANPASPPLSLVLRHLCEHFRVLSVYTHSSSVKSPENLLKCTGEONKKQPROD 180  
QY 181 YOLGFTLLIMKNVPTKQMKFSIQTMCPTEGEGNIARFLFSLFGQKNAVNAATLIDSWDIA 240  
DB 181 YOLGFTLLIMKNVPTKQMKFSIQTMCPTEGEGNIARFLFSLFGQKNAVNAATLIDSWDIA 240  
QY 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVILMSVLOQIGGCSVTVPANVQ 300  
DB 241 IFOLKEGSSKEKAAVFRSMNSALGKSPMLAGNELTVADVILMSVLOQIGGCSVTVPANVQ 300  
QY 301 RWRMSCENLAPFNLTALKLK 320  
DB 301 RWRMSCENLAPFNLTALKLK 320

```
RESULT 3
Q8R010 PRELIMINARY; PRT; 320 AA.
ID Q8R010
AC Q8R010;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Jtvl-pending protein.
GN Name-Jtvl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N, and FVB/N-3; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024410; AAH24410.1; -.
DR EMBL; BC026972; AAH26972.1; -.
DR MGI; MGI:2385237; Jtvl.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
SQ SEQUENCE 320 AA; 35396 MW; 1745D7EB4BC3670D CRC64;

Query Match 87.8%; Score 1464; DB 2; Length 320;
Best Local Similarity 87.2%; Pred. No. 1.8e-110;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEMYQVPRYHGGAPRLVELPTCMYRLPVNHGRSYGPAGAGHVOEESNLSIQALESROD 60
DB 1 MEMYQVPRYHGGAPRLVELPTCMYRLPVNHGRSYGPAGAGHVOEESNLSIQALESROD 60
QY 61 DILKRLYLKAAVVDGSKMIHTPPDADLDVTNIIQADEPTTLTNTLADLSVGLKDYGALK 120
DB 61 DILKRLYLKAAVVDGSKMIHTPPDADLDVTNIIQADEPTTLTNTLADLSVGLKDYGALK 120
QY 121 DIVINANPASPLSLIVLHRLCEHFRVLTSTVHTSSVKSVPENLLKCGEONKQKPROD 180
DB 121 DIVINANPASPLSLIVLHRLCEHFRVLTSTVHTSSVKSVPENLLKCGEONKQKPROD 180
QY 181 YOLGFTLLWKNVPTQMKFESVQTMCEPIEGEGNIARFLFSLFGQKHNAVATLLIDSVDIA 240
DB 181 YOLGFTLLWKNVPTQMKFESVQTMCEPIEGEGNIARFLFSLFGQKHNAVATLLIDSVDIA 240
```

```
DB 181 YOLGFTLLWKNVPTQMKFESVQTMCEPIEGEGNIARFLFSLFGQKHNAVATLLIDSVDIA 240
QY 241 IFOLKESKSKERAAVRSNNSALGSKPMLAGNELTVADVLNLSVLIQIGCSVTVPANVO 300
DB 241 IFOLKESKSKERAAVRSNNSALGSKPMLAGNELTVADVLNLSVLIQIGCSVTVPANVO 300
QY 301 RMRWSCENLAPNTLAKLKLK 320
DB 301 RMRWSCENLAPNTLAKLKLK 320

RESULT 4
Q8R2Y6 PRELIMINARY; PRT; 320 AA.
ID Q8R2Y6
AC Q8R2Y6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Jtvl-pending protein.
GN Name-Jtvl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026958; AAH26958.1; -.
DR MGI; MGI:2385237; Jtvl.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
SQ SEQUENCE 320 AA; 35423 MW; 1C21FLA74C9882B4 CRC64;

Query Match 87.8%; Score 1464; DB 2; Length 320;
Best Local Similarity 87.2%; Pred. No. 1.8e-110;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEMYQVPRYHGGAPRLVELPTCMYRLPVNHGRSYGPAGAGHVOEESNLSIQALESROD 60
DB 1 MEMYQVPRYHGGAPRLVELPTCMYRLPVNHGRSYGPAGAGHVOEESNLSIQALESROD 60
QY 61 DILKRLYLKAAVVDGSKMIHTPPDADLDVTNIIQADEPTTLTNTLADLSVGLKDYGALK 120
DB 61 DILKRLYLKAAVVDGSKMIHTPPDADLDVTNIIQADEPTTLTNTLADLSVGLKDYGALK 120
QY 121 DIVINANPASPLSLIVLHRLCEHFRVLTSTVHTSSVKSVPENLLKCGEONKQKPROD 180
DB 121 DIVINANPASPLSLIVLHRLCEHFRVLTSTVHTSSVKSVPENLLKCGEONKQKPROD 180
```

```

Db      121 DIVINANPASPPLSLVILVHRLLCERVSLVTHSHSVKNVENVLCFGEQARQSRHE 180
Oy      181 YOLGFTLIMKNVPTQMKFSIQTMCPPIEGEGNIARFLPSLFGQKHNANVATLIDSWDIA 240
Db      181 YOLGFTLIMKNVPTQMKFSVQTMCPPIEGEGNIARFLPSLFGQKHNANVATLIDSWDIA 240
Oy      241 IFOLKEGSSKEXKAAVFRSMNSALGKSPMLAGNELTVADVLMVSLVLOQIGGCVTVPANVQ 300
Db      241 MFOLKEGSSKEXKAAVFRSMNSALGKSPMLVGNELTVADVLMVSLVLOQIGGSGAAPTANVQ 300
Oy      301 RMWRCENLAPFPTALKLK 320
Db      301 RMWRCENLAPFPTALKLK 320

```

```

RESULT 5
MCA2_CRIGR STANDARD; PRT; 320 AA.
AC 09WVW7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Multisynthetase complex auxiliary component p38.
OS Criceuluss griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=99096915; PubMed=9878398; DOI=10.1006/jmbi.1998.2316;
RA Quevillon S., Robinson J.-C., Berthomieu E., Siatecka M., Mirande M.;
RT "Macromolecular assemblage of aminoacyl-tRNA synthetases:
RT identification of protein-protein interactions and characterization of
RT a core protein."
RL J. Mol. Biol. 285:183-195(1999).

```

```

CC -1- FUNCTION: Probable core protein of the multisynthetase complex
CC that serves as a template for the assembly of the supramolecular
CC structure.
CC -1- SUBUNIT: Component of the multisynthetase complex which is
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,
CC arginyl, and asparagyl-tRNA synthetases as well as three auxiliary
CC proteins, p18, p48 and p43.
CC -1- SIMILARITY: Contains 1 GST-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF072727; AAD38422.1; -
DR InterPro: IPR010987; GST_C_1like.
DR InterPro: IPR004046; GST_C-term.
DR Pfam: PF00043; GST_C; 1.
KW Protein biosynthesis.
SQ
SEQUENCE 320 AA; 35433 MW; 6D24E033ABEC10A CRC64;

```

```

Query Match      86.6%; Score 1444; DB 1; Length 320;
Best Local Similarity 86.6%; Pred. No. 7.6e-109;
Matches 277; Conservative 16; Mismatches 27; Indels 0; Gaps 0;

```

```

Oy      1 MEMYGVKPYHGGAGLRYELPTCMYRLPNVHGSGVPGAFGAGHVOESNLSLOALSSROD 60
Db      1 MEMYGVKSYHGGASRLRYELPTCMYRLPNVHSGKTSPTADAGHVOETSEPSLOALSSROD 60
Oy      61 DILRLYLKAAVDSGSKMIDPADLDVNTIIQADEPTTLTNALDINSLVGLKQYALK 120

```

```

Db      61 DILRLYLKAAVDSGSKMIDPADLDVNTIIQADEPTTLTNALDINSLVGLKQYALK 120
Oy      121 DIVINANPASPPLSLVILVHRLLCERFVSLVTHSHSVSNVSEBENILKCGEGNKQOPROD 180
Db      121 DIVINANPASPPLSLVILVHRLLCERFVSLVTHSHSVKNVENVLCFGEQARQSRHE 180
Oy      181 YOLGFTLIMKNVPTQMKFSIQTMCPPIEGEGNIARFLPSLFGQKHNANVATLIDSWDIA 240
Db      181 YOLGFTLIMKNVPTQMKFSVQTMCPPIEGEGNIARFLPSLFGQKHNANVATLIDSWDIA 240
Oy      241 IFOLKEGSSKEXKAAVFRSMNSALGKSPMLAGNELTVADVLMVSLVLOQIGGCVTVPANVQ 300
Db      241 MFOLKEGSSKEXKAAVFRSMNSALGKSPMLVGNELTVADVLMVSLVLOQIGGSGAAPTANVQ 300
Oy      301 RMWRCENLAPFPTALKLK 320
Db      301 RMWRCENLAPFPTALKLK 320

```

```

RESULT 6
Q8R3V2 PRELIMINARY; PRT; 280 AA.
ID Q8R3V2;
AC Q8R3V2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE JTV1.
GN Name=Jtv1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheef C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hejblum F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lounellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fanev J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerich A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC024480; AAH24480.1; -
DR MGD: MGI:2385237; JTV1.
DR InterPro: IPR004046; GST_C-term.
DR InterPro: IPR010987; GST_C_1like.
DR Pfam: PF00043; GST_C; 1.
SQ
SEQUENCE 280 AA; 31097 MW; E5BE81498983FD2E CRC64;

```

```

Query Match      75.6%; Score 1260; DB 2; Length 280;
Best Local Similarity 88.0%; Pred. No. 5.5e-94;
Matches 243; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

```

```

Oy      45 Q8ESNLSIQALSSRODILRLYLKAAVDSGSKMIDPADLDVNTIIQADEPTTLTN 104

```

Db	5	QETSEPSLQALBESRODILKRLYLEKQADGLSKMHTHPADLDVNNILQADEPTLATN	64
Qy	105	ALDINSVLKQYDYGALKOIYINANPASPPLSLYLHRLCEHFVLSLWTHSSVKSYPEN	164
Db	65	TLDLSVLGKDYGALKOIYINANPASPPLSLYLHRLCGRVLSLWTHSSVKYENPEN	124
Qy	165	ILKCFGEQNKQPRODYDQFTLIMKNVPTKQMKFSIQTCPIEGEKNIRAPLFSLEQOK	224
Db	125	LYKCFGEQAKRQSRHEYOGLFTLIMKNVPTKQMKFSQVOTCPIEGEKNIRAPLFSLEQOK	184
Qy	225	HNAVNATLDSWVDIAI POLKEGSSKEKA VPFSSMSALQKSPMLACNGLTVADVLMV	284
Db	185	HNAVUTLTDLSWVDIMFQLRBESSKEKAVPFSSMSALORSWMLVSNELTVADVLMV	244
Qy	285	LQQTGGCSVTVPANVORMRSCNLPFNTATKLK	320
Db	245	LQQTGGSSGAAPTNVQRMKSCNLPFPSTALDK	280

Query Match	Score 114.5	DB 2	Length 311
Best Local Similarity	68.8%	Pred. No. 1.5e-84	
Matches 220	Conservative 43	Mismatches 48	Indels 9
			Gaps 3

[illegible]

ID	Q6IN04	PRELIMINARY;	PRT;	311 AA.
Q6IN04				
AC	Q6IN04;			
DT	05-JUL-2004 (TREMBLrel. 27, Created)			
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)			
DE	MG680304 protein.			
GN	Name=MG680304;			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=22368257; PubMed=12477932; DOI=10.1073/pnas.242603699;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh P.,			
RA	Diatchenko L., Marisusa K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Ueff T.B., Tshiyuki S., Carrini P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Foley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whitting M., Madan A., Young A.C., Shchepochko Y., Bouffard G.G.,			
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodríguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywicki M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,			
RA	Jones S.J., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;			
RA	Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,			
RA	Richardson P.;			
RT	"Genetic and genomic tools for Xenopus research: The NIH Xenopus			
RT	initiative."			
RL	Dev. Dyn. 225:384-391(2002).			
RN	[3]			

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RA Klein S., Gerhard D.S.;

RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC072178; AAH72178.1;

DR InterPro: IPR004046; GST\_Cterm.

DR InterPro: IPR010987; GST\_C\_1like.

DR Pfam: PF00043; GST\_C\_1.

DR S0 SEQUENCE 311 AA; 34424 MW; D5E8325C18D86751 CRC64;

Query Match 68.4%; Score 1139.5; DB 2; Length 311;

Best Local Similarity 69.1%; Pred. No. 3.8e-84;

Matches 221; Conservative 40; Mismatches 50; Indels 9; Gaps 3;

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Klein S.L., Strassberg R.L., Wagner L., Pontius J., Clifton S.W.,

RA Richardson P.;

RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus

RA Dev. Dyn. 225:384-391(2002).

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Klein S., Strassberg R.;

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC043832; AAH43832.1;

DR InterPro: IPR004046; GST\_Cterm.

DR InterPro: IPR010987; GST\_C\_1like.

DR Pfam: PF00043; GST\_C\_1.

DR S0 SEQUENCE 311 AA; 34647 MW; 2F08C9DA60DD63BE CRC64;

Query Match 67.9%; Score 1132.5; DB 2; Length 311;

Best Local Similarity 69.1%; Pred. No. 1.4e-83;

Matches 221; Conservative 39; Mismatches 51; Indels 9; Gaps 4;

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

1 MEMYQVPPYHGGAPLRYELPTCMYRLPNVHSGYGPAPAGHVEESNLSIQLESROD 60

RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Hasteh F.  
RA Diatchenko L., Marnusina K., Farmer A.A., Rubin G.M., Hong L.,  
R Scapleton M., Soares M.B., Bonaldo M.P., Cavaavant I.L., Schetz Z.E.,  
RA Brownstein N.J., Ussin T.B., Toohyuk S., Carinci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Boask S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulky S.W.,  
RA Villalon D.K., Muzy D.N., Sodergren E.U., Lu X., Gibbs R.A.,  
RA Fahney J., Helton B., Ketteman M., Madam A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madam A.C., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J.J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalios D.E., Schmercn A., Schein J.E.,  
RA Jones S.J., Maria M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Kidney;  
RC Strubeberg R.;  
RL Submitted (JUN-2003) to the EMBL/genbank/DDBJ databases.  
DR EMBL; BC053178; AAH53178.1; -;  
DR ZFIN; ZDB-GENE-040426-2652; zgs:63976.  
DR InterPro; IPRO04046; GST\_Cterm.  
DR Interpro; IPRO10987; GST\_C\_Like.  
DR Pfam; PF00043; GST\_C\_1  
SQ SEQUENCE 321 bp; 34852 MW; BABF6B951208244A CRC64;

	Query Match	53.2%;	Score 887.5;	DB 2;	length 321;	
	Best Local Similarity	55.6%;	Pred. No. 1,le-63;			
	Matches	184;	Conservative	50;	Mismatches	76;
					Indels	21;
					Gaps	8

  

Oy	1	MPMTQVKEHYGGAGLRYVELPTCMYRLFNHGKSYGPAAPGAHQVE--ESNLSLOALRESRQ	59
Dd	1	MPMYOVKEV--SPADITVDLPCTMKYLEPNVAHQ--GASLGHEHALONGEWDPPVKALESRQ	56
Oy	60	DDLRKLRELKAAYVGLSKMIQTDPADDVTNNIQADBPTTLT-----TNALDMSV	111
Dd	57	DELRKLYELKATVDGLAKTYTTTPPADDDASTLAH-----TTHTPPALAVLRGTADLDL	111
Oy	112	LKGQDALKDIVINANPASPILLSLVLRHLCEHFVLSTVHTSHSVKSVPENILCKFGE	171
Dd	112	LKGQDALGRDIVINANPAQPPLSLIYLTHALLCQRFOQLSSVHHSVSFTVPAPFLSCGP	171
Oy	172	ON-KKQRPQDVLTGFTLLTMKNVKTOMKFISOTMCPRIEGGNIAEFSLSG-QKHNAVN	229
Dd	172	RHTHSYAHRRPOLGTLTIWKDSYKIQMFSTONMCPRIEGGVNAFFLYRLIGAEPRDPS	231
Oy	230	ATLIDSWVDIAIFOLKEGSSKEKAIVFRSMNSALGKSPMLAGNELITVADVLMVSLQQIG	289
Dd	232	ATLMGWDTALFOLAEGSKERRAAVLRALNALGRSPMLIGQEPSLADIYSACCVLTQG	291
Oy	290	GCSYTVPANQRMMSCENTLAPFNALCLK	320
Dd	292	QTS-SAPANVORWLKSCNLGYFSVCVDPILQ	321

  

RESULT	11	
ID	0707A3	PRELIMINARY; PRT; 340 AA.
NC	0707A3:	
DT	01-MAR-2004 (TREMBLrel. 26, Created)	
DT	01-MAR-2004 (TREMBLrel. 26, Last sequence update)	
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)	
DE	AGCF5808 (Fragment).	
GN	Name=agCG50514; ORFNames=ENSNANGG00000011827;	
OS	Anopheles gambiae str. PEEST.	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Psatygota;	
CC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.	
OX	NCBTaxid=180454;	
NP	(1)	
	SEQUENCE FROM N.A.	

RC STRAIN=PEST;  
RA Anopheles Genome Sequencing Consortium;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data. EAA10870.1; -.  
DR EMBL; AAB01008960;  
FT NON-ITER 1  
SQ SEQUENCE 340 AA; 37467 MW; FD0DB525CDE4B976 CRC64;

Query Match 100; Conservative 47; Mismatches 13; Indels 63; Gaps 13;  
16.0%; Score 267.5; DB 2; Length 340;  
Best Local Similarity 29.2%; Pred. No. 2,66-13;  
Matches 100; Conservative 47; Mismatches 13; Indels 63; Gaps 13;





```

QY      183  GGFTLLMKNVKTOMKPSIQTCMPLEEGSGNARPLFSLFGOKHNAVNTL---IDSWDI 239
           |||||
DB       180  ISVTLLMNCKEHTEMISPTTWVPFLYEENVIIRYGRVGPAERYEGSPLCNEIDLVDI 239
           |||||
QY       240  AIFOLKESSSEKAAVFRSMNSALCKSPMLAGNELTVADVVLMSTVLQDIGCSVTVPANV 299
           |||||
DB        240  CYQLLRCHTHTKTQVAMVRLLDRLRLOKKQYFGGSCMSVADVGYSSL-----TRMPAVT 292
           |||||
QY        300  QR 301
           ::
DB         293  EK 294
           ::

RESULT 14
Q6NMK4    PRELIMINARY; PRT; 322 AA.
O6NMK4
AC O6NMK4;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Lp12114p (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stepleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RL Submitted (May-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BT012670; AA08476.1; -.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1-like.
PF Pfam; PF00043; GST_C_1.
FT NON_TER
SQ SEQUENCE 322 AA; 35528 MW; 82A15927DF928DAE CRC64;

Query Match 13.4%; Score 224; DB 2; Length 322;
Best Local Similarity 28.1%; Pred. No. 8,3e-10;
Matches 85; Conservative 47; Mismatches 120; Indels 50; Gaps 10.

QY      18  VELPTCMYRLPNV-----HGNSYGAPGAGHVQESNISLQLESRODDILKRYEL 69
           |||||
DB       34  IKLPCTCWPPLKNVSLADSLASGSSTSASTSSTSCDDTASVA--RQEKVLRKLDEL 90
           |||||
QY       70  KAAVDGLSKMTOTPRADDYNNIIADDEPTTLTTNALDLNSTLGGDYGLAKDIYINANPA 129
           |||||
DB        91  KAOLQAIR-----AGLGVCG-KTFPHTTAFORG-----GIKEVPLDDVINGHFN 134
           |||||
QY       130  SPRLSLVLNRLLSCEHFVLTSTVTHSSSVKSY-----PENLRKCFGEONKKOPROXYQ 182
           |||||
DB        135  FIVAYLLDLKANMRVLYITIDVKTFTSHSTVADI GPAREFEANLAKV--PVNPALPK---- 188
           |||||
QY       183  LGFTLLMKNVKPTQMKFSIQTMCPIREGSGNIARPLFSLFQGNANVANNTL---IDSWDI 239
           |||||
DB       189  ISVTLLMNCKEHTEMISPTTWVPFLYEENVIIRYGRVGPAERYEGSPLCNEIDLVDI 248
           |||||
QY       240  AIFOLKESSSEKAAVFRSMNSALCKSPMLAGNELTVADVVLMSTVLQDIGCSVTVPANV 299
           |||||
DB        249  CYQLLRCHTHTKTQVAMVRLLDRLRLOKKQYFGGSCMSVADVGYSSL-----TRMPAVT 301
           |||||
QY        300  OR 301
           ::
DB         302  EK 303
           ::

RESULT 15
MCA2_DROME STANDARD; PRT; 334 AA.
AC O9VUR3;
DT 16-OCT-2001 (rel. 40; Created)
```

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 25-JUN-2005 (Rel. 46, Last annotation update)  
DE Probable multisynthetase complex auxiliary component p38.  
GN ORFNames=CG12304;  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephydroidae; Drosophilidae; Drosophila.  
CX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkelley;  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celinkier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aggalyan A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Broctier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pallos B., Delcher A., Deng Z., Duan A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fowler C., Gabriellian A.E., Garg N.S., Galbati W.M., Glasser K.,  
RA Glodde J., Gong F., Gorelli J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Hargreave D.A., Heitman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajalil M., Kalish F., Kapten G.H., Ke Z., Kemsinn J.A., Ketchum K.A.,  
RA Kimmel B.E., Koditz C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lian X., Lai Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mateci B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Mialshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Patel J.M.,  
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E.B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitekbae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RL "The genome sequence of *Drosophila melanogaster*.";  
RT Science 287:2185-2195(2000).  
RN [2]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Hradek S., Crosby W.A., Wungkil C.U., Matthews B.B., Campbell K.S.,  
RA Midszky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celinkier S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.B.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RL systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
CC -I- FUNCTION: Probable core protein of the multisynthetase complex  
CC that serves as a template for the assembly of the supramolecular  
CC structure (By similarity).  
CC -I- SUBUNIT: Component of the multisynthetase complex which is  
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the  
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary  
CC proteins, p18, p48 and p43 (By similarity).  
CC -I- SIMILARITY: Contains 1 GST-like domain.



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 23, 2005, 13:25:39 ; Search time 70.371 Seconds  
(without alignments)  
1758.725 Million cell updates/sec

Title: US-10-622-817-3

Perfect score: 1667

Sequence: 1 MPWGVKRVHSGAPLRVRL.....RWRSCENLAPRTALKLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneeqp19808:\*  
2: geneeqp19908:\*  
3: geneeqp20008:\*  
4: geneeqp20018:\*  
5: geneeqp20028:\*  
6: geneeqp20038:\*  
7: geneeqp20038:\*  
8: geneeqp20048:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	100.0	320	7	ADE58613 Human Pro
2	1667	100.0	320	8	ABM80843 Tumour-as
3	1630	97.8	312	2	AAW25776 UTU1 prot
4	1630	97.8	312	8	ADR86551 1-312 am
5	1378	82.7	272	7	ADC10204 Human NOV
6	1191	71.4	229	8	ADR86553 84-312 am
7	832	49.9	161	8	ADR86552 1-161 am
8	247	14.8	51	5	ABP01502 Human ORF
9	214.5	12.9	334	4	ABE82468 Drosophila
10	181	10.9	263	4	ABG23964 Novel hum
11	125.5	7.5	716	3	AAQ1411 Arabidops
12	125.5	7.5	719	3	AAQ1410 Arabidops
13	125.5	7.5	748	3	AAQ1409 Arabidops
14	104	6.2	639	4	ABE82218 Drosophila
15	103	6.2	1512	4	AAU04349 Mammalian
16	103	6.2	1512	4	AAU04349 Mammalian
17	103	6.2	1550	4	AAU04349 Mammalian
18	102	6.1	201	8	AD824684 Bacterial
19	101	6.1	559	2	AAAR10682 Polynidro
20	101	6.1	826	4	AAAR10682 Polynidro
21	101	6.1	826	4	AAAR10682 Polynidro
22	101	6.1	826	4	AAAR10682 Polynidro
23	101	6.1	833	4	AAAR10682 Polynidro
24	101	6.1	833	4	AAAR10682 Polynidro
25	101	6.1	833	4	AAAR10682 Polynidro

26	101	6.1	1221	4	AAQ1409 C. glutam
27	101	6.1	1221	5	ABG80321 C. glutam
28	101	6.1	1221	5	AAE18908 Coryneb
29	101	6.1	1221	7	ADD13589 C. glut
30	100.5	6.0	1200	6	ABU26059 Protein e
31	99.5	6.0	650	2	AAI29614 Mouse PIA
32	99.5	6.0	651	2	AAI29621 Mouse PIA
33	97.5	5.8	217	7	ADBE3092 Rat Prote
34	97.5	5.8	217	7	ADDA8150 Rat Prote
35	97.5	5.8	217	7	ADDA8158 Rat Prote
36	97.5	5.8	217	7	ADDA8154 Rat Prote
37	97.5	5.8	217	7	ADBE3084 Rat Prote
38	97.5	5.8	217	7	ADBE3088 Rat Prote
39	97.5	5.8	218	2	AAI20033 Rat brain
40	97.5	5.8	218	7	ADBE5502 Rat Prote
41	97.5	5.8	652	5	ABP41869 Human ova
42	97	5.8	559	3	AAI77178 S. venezu
43	97	5.8	2221	4	ABE61339 Drosophila
44	97	5.8	5215	8	ADL91930 Streptomy
45	96.5	5.8	1049	6	ABM67090 Photorhab

## ALIGNMENTS

RESULT 1	ADBE58613	standard; protein; 320 AA.
ID	ADBE58613	standard; protein; 320 AA.
AC	ADBE58613	
DT	29-JAN-2004	(first entry)
DE	Human Protein Q13155, SEQ ID NO 4489.	
XX	Human; pain; neuronal tissue; gene therapy;	
XX	spinal segmental nerve injury; chronic constriction injury; CCI;	
XX	spared nerve injury; SNI; Chung.	
OS	Homo sapiens.	
XX	WO2003016475-A2.	
XX	27-FEB-2003.	
XX	14-AUG-2002; 2002WO-US025765.	
XX	14-AUG-2001; 2001US-0312147P.	
XX	01-NOV-2001; 2001US-0346382P.	
XX	26-NOV-2001; 2001US-033347P.	
XX	(GENO) GEN HOSPITAL CORP.	
XX	(FARB) BAYER AG.	
XX	Woolf C, D'ureo D, Befort K, Costigan M;	
XX	WPI, 2003-268312/26.	
XX	GENBANK; Q13155.	
XX	New composition comprising two or more isolated polypeptides, useful for	
XX	preparing a medicament for treating pain in an animal.	
XX	Claim 1, Page; 1017pp; English.	
XX	The invention discloses a composition comprising two or more isolated rat	
XX	or human polynucleotides or a polynucleotide which represents a fragment,	
XX	derivative or allelic variation of the nucleic acid sequence. Also	
XX	claimed are a vector comprising the novel polynucleotide, a host cell	
XX	comprising the vector, a method for identifying a nucleotide sequence	
XX	which is differentially regulated in an animal subjected to pain and a	
XX	kit to perform the method, an array, a method for identifying an agent	
XX	that increases or decreases the expression of the polynucleotide sequence	
XX	that is differentially expressed in neuronal tissue of a first animal	

CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 320 AA;

Query Match 100.0%; Score 1667; DB 7; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-167;  
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYQVYPYHGGAPLRLVELPTCMYRLPNVHGRSGYPAGAGHVOEBSNLSLOLESROD 60  
 DB 1 MPWYQVYPYHGGAPLRLVELPTCMYRLPNVHGRSGYPAGAGHVOEBSNLSLOLESROD 60  
 QY 61 DILKRLYELKAAVVDGSKMLOTDPADLDVTNIIQADEPTTLTTNALDLSVLTGKDYGALK 120  
 DB 61 DILKRLYELKAAVVDGSKMLOTDPADLDVTNIIQADEPTTLTTNALDLSVLTGKDYGALK 120  
 QY 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPENLLKCFEBOKKOPROD 180  
 DB 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPENLLKCFEBOKKOPROD 180  
 QY 181 YOLGFTLLTWKNVPTKQKFSIQTCPIEGEGNIAFLFSLFGQKHNAVNATLIDSWVDA 240  
 DB 181 YOLGFTLLTWKNVPTKQKFSIQTCPIEGEGNIAFLFSLFGQKHNAVNATLIDSWVDA 240  
 QY 241 IFOLKEGSSKEKAIVFRSMNSALGKSPWLAGNELTVADVLMVSVLQOIGGCVTPVAVNQ 300  
 DB 241 IFOLKEGSSKEKAIVFRSMNSALGKSPWLAGNELTVADVLMVSVLQOIGGCVTPVAVNQ 300  
 QY 301 RWMRSCEMLAPFNTALTKLK 320  
 DB 301 RWMRSCEMLAPFNTALTKLK 320

RESULT 2

ABM80843 standard; protein; 320 AA.

XX ABM80843;

DT 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) polypeptide PRO81501, SEQ:2179.

KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KW cervical cancer; bladder cancer; pancreatic cancer;  
 KW chromosomal identification; leukaemia; hybridisation probe;  
 KW gene therapy; cytostatic.

XX Homo sapiens.

OS WO2004030615-A2.

XX 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.  
 PF 02-OCT-2002; 2002US-0414971P.  
 PR (GENTH) GENENTECH INC.  
 PA Wu TD, Zhang Z, Zhou Y;  
 PI WPI; 2004-347921/32.  
 DR N-PSDB; ACN38579.  
 XX New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 PS Claim 12; SEQ ID NO 2179; 7273bp; English.  
 XX The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acid and polypeptide  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT polypeptide of the invention.

XX Sequence 320 AA;

Query Match 100.0%; Score 1667; DB 8; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-167;  
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYQVYPYHGGAPLRLVELPTCMYRLPNVHGRSGYPAGAGHVOEBSNLSLOLESROD 60  
 DB 1 MPWYQVYPYHGGAPLRLVELPTCMYRLPNVHGRSGYPAGAGHVOEBSNLSLOLESROD 60  
 QY 61 DILKRLYELKAAVVDGSKMLOTDPADLDVTNIIQADEPTTLTTNALDLSVLTGKDYGALK 120  
 DB 61 DILKRLYELKAAVVDGSKMLOTDPADLDVTNIIQADEPTTLTTNALDLSVLTGKDYGALK 120  
 QY 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPENLLKCFEBOKKOPROD 180  
 DB 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPENLLKCFEBOKKOPROD 180  
 QY 181 YOLGFTLLTWKNVPTKQKFSIQTCPIEGEGNIAFLFSLFGQKHNAVNATLIDSWVDA 240  
 DB 181 YOLGFTLLTWKNVPTKQKFSIQTCPIEGEGNIAFLFSLFGQKHNAVNATLIDSWVDA 240  
 QY 241 IFOLKEGSSKEKAIVFRSMNSALGKSPWLAGNELTVADVLMVSVLQOIGGCVTPVAVNQ 300  
 DB 241 IFOLKEGSSKEKAIVFRSMNSALGKSPWLAGNELTVADVLMVSVLQOIGGCVTPVAVNQ 300  
 QY 301 RWMRSCEMLAPFNTALTKLK 320  
 DB 301 RWMRSCEMLAPFNTALTKLK 320

RESULT 3

AAW25776

ID AAW25776 standard; protein; 312 AA.  
XX  
AC AAW25776;  
XX  
DT 19-DEC-1997 (first entry)  
XX  
XX JTV1 protein.  
DE  
XX JTV1; hPMS2; probe; detection; chromosome 7; deletion;  
KM mismatch repair gene; hereditary non-polyposis colorectal cancer;  
KM homologous recombination.  
XX  
OS Homo sapiens.  
XX  
PN MO9708312-A1.  
XX  
PD 06-MAR-1997.  
XX  
PF 26-AUG-1996; 96MO-US013598.  
XX  
PR 24-AUG-1995; 95US-00518862.  
XX  
PA (UYUO ) UNIV JOHNS HOPKINS.  
XX  
PI Vogelstein B, Kinzler KW, Nicolaides NC;  
XX  
PI MPI; 1997-179269/16.  
DR N-PSDB; AAT86182.  
XX  
PT Novel chromosome 7 gene, JTV1 - used for detecting chromosome 7  
XX deletions, and PMS2 promoter activity.  
XX  
XX Claim 5; Fig 2; 31pp; English.  
XX  
XX This sequence is JTV1 protein and is encoded by DNA isolated from human  
XX chromosome 7. The JTV1 coding sequence is located upstream from hPMS2.  
XX JTV1 cDNA can be used as probes to detect chromosome 7 deletions  
XX involving JTV1. Due to the overlapping promoter regions, deletions of  
XX JTV1 would also affect PMS2 (a mismatch repair gene) expression, leading  
XX to hereditary non-polyposis colorectal cancer. JTV1 can also be used to  
XX assay activity or competence of the PMS2 promoter region, the presence of  
XX JTV1 suggesting that the PMS2 promoter is intact. JTV1 sequences can also  
XX be used to guide homologous recombination at the PMS2 locus  
XX  
XX Sequence 312 AA;  
SQ

Query Match 97.8%; Score 1630; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPMYQVKPYPHGGAPLRVLPCTMYRLPNVHGSRYPAGAGHVOEESNLSLOALSSROD 60  
DB 1 MPMYQVKPYPHGGAPLRVLPCTMYRLPNVHGSRYPAGAGHVOEESNLSLOALSSROD 60  
QY 61 DIKRLYEYKAAVDSKMIQTPDADLDVTNIIQADEPTLTNNALDLSVLGKDYGALK 120  
DB 61 DIKRLYEYKAAVDSKMIQTPDADLDVTNIIQADEPTLTNNALDLSVLGKDYGALK 120  
QY 121 DIVINNAPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLLKCFGEQNKQPROD 180  
DB 121 DIVINNAPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLLKCFGEQNKQPROD 180  
QY 181 YOLGFTLIMKNVKTOMKRSIQTMCPTEBEGNIAFLFSLFGQKNAVNATLLDSWVDA 240  
DB 181 YOLGFTLIMKNVKTOMKRSIQTMCPTEBEGNIAFLFSLFGQKNAVNATLLDSWVDA 240  
QY 241 IFOLKGGSSKEKAIVRRSMNSALGKSPWLAGNELTYADVVLMSVLOIGGCVTVFANYO 300  
DB 241 IFOLKGGSSKEKAIVRRSMNSALGKSPWLAGNELTYADVVLMSVLOIGGCVTVFANYO 300  
QY 301 RWRMSCENTLAPF 312  
DB 301 RWRMSCENTLAPF 312

RESULT 4  
ID ADR86551  
ID ADR86551 standard; protein; 312 AA.  
XX  
AC ADR86551;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE 1-312 amino acid sequence of p38/JTV-1 protein.  
XX  
KM p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.  
XX  
OS Homo sapiens.  
XX  
PN EP1454628-A2.  
XX  
PD 08-SEP-2004.  
XX  
PF 09-SEP-2003; 2003EP-00020344.  
XX  
PR 03-MAR-2003; 2003KR-00013058.  
XX  
PA (UYSE-) UNIV SEOUL NAT IND FOUND.  
XX  
PI Kim S, Park B;  
XX  
PI MPI; 2004-627822/61.  
DR N-PSDB; ADR86548.  
XX  
PT New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
XX e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
XX and cutaneous or intraocular melanoma, as well as for screening new  
XX anticancer agents.  
XX  
XX Claim 5; SEQ ID NO 4; 47pp; English.  
XX  
XX The present invention relates to an isolated p38/JTV-1 protein for use as  
XX medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
XX useful as medicament for treating breast cancer, large intestinal cancer,  
XX lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
XX cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
XX cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
XX rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
XX endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
XX Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
XX cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
XX tumour, urethral cancer, penile cancer, prostate cancer, chronic or acute  
XX leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
XX cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
XX CNS lymphoma, bone marrow tumour, brain stem nerve glioma, pituitary  
XX adenoma, or their combination. The protein is useful as a target for  
XX screening new anticancer agents. The present sequence represents the 1-  
XX 312 amino acid sequence of p38/JTV-1 protein.  
XX  
XX Sequence 312 AA;  
SQ

Query Match 97.8%; Score 1630; DB 8; Length 312;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPMYQVKPYPHGGAPLRVLPCTMYRLPNVHGSRYPAGAGHVOEESNLSLOALSSROD 60  
DB 1 MPMYQVKPYPHGGAPLRVLPCTMYRLPNVHGSRYPAGAGHVOEESNLSLOALSSROD 60  
QY 61 DIKRLYEYKAAVDSKMIQTPDADLDVTNIIQADEPTLTNNALDLSVLGKDYGALK 120  
DB 61 DIKRLYEYKAAVDSKMIQTPDADLDVTNIIQADEPTLTNNALDLSVLGKDYGALK 120  
QY 121 DIVINNAPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLLKCFGEQNKQPROD 180  
DB 121 DIVINNAPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLLKCFGEQNKQPROD 180

QY 181 YOLGFTLLIKWNVKPTOMKFSIQTMCPTEGEGNIARFLFSLFGQHNANVATLIDSWDIA 240  
 XX |||||||  
 PA 181 YOLGFTLLIKWNVKPTOMKFSIQTMCPTEGEGNIARFLFSLFGQHNANVATLIDSWDIA 240  
 QY 241 IFOLKEGSSKEKAAYFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCSYTVPANVQ 300  
 XX |||||||  
 DB 241 IFOLKEGSSKEKAAYFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCSYTVPANVQ 300  
 QY 301 RMRSCENLAPF 312  
 XX |||||||  
 DB 301 RMRSCENLAPF 312  
 RESULT 5  
 ADCL0204  
 ID ADC10204 standard; protein; 272 AA.  
 AC ADC10204;  
 XX  
 DT 18-DEC-2003 (first entry)  
 DE Human NOVX polypeptide SEQ ID NO: 226.  
 XX  
 KW cytosolic; antidiabetic; anorectic; cerebroprotective; neuroprotective;  
 KW antiinflammatory; gene therapy; antisense therapy; thymomimetic; NOVX;  
 KW pathology; cancer; diabetes; obesity; endocrine disorder; CNS disorder;  
 KW inflammatory disorder; chromosome mapping; tissue typing;  
 KW predictive medicine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003000842-A2.  
 XX  
 PD 03-JUN-2003.  
 PF 04-JUN-2002; 2002WO-US017443.  
 XX  
 PR 04-JUN-2001; 2001US-0295607P.  
 XX  
 PR 04-JUN-2001; 2001US-0295661P.  
 PR 06-JUN-2001; 2001US-0296404P.  
 PR 06-JUN-2001; 2001US-0296418P.  
 PR 07-JUN-2001; 2001US-0296575P.  
 PR 11-JUN-2001; 2001US-0297414P.  
 PR 12-JUN-2001; 2001US-0295573P.  
 PR 12-JUN-2001; 2001US-0297567P.  
 PR 14-JUN-2001; 2001US-0298285P.  
 PR 15-JUN-2001; 2001US-0298528P.  
 PR 18-JUN-2001; 2001US-0299133P.  
 PR 19-JUN-2001; 2001US-0299230P.  
 PR 21-JUN-2001; 2001US-0299949P.  
 PR 22-JUN-2001; 2001US-0300177P.  
 PR 26-JUN-2001; 2001US-0300883P.  
 PR 28-JUN-2001; 2001US-0301530P.  
 PR 28-JUN-2001; 2001US-0301550P.  
 PR 03-JUL-2001; 2001US-0302951P.  
 PR 31-JUL-2001; 2001US-0308890P.  
 PR 14-SEP-2001; 2001US-0322297P.  
 PR 25-SEP-2001; 2001US-0324669P.  
 PR 03-DEC-2001; 2001US-0337477P.  
 PR 14-DEC-2001; 2001US-0341562P.  
 PR 21-FEB-2002; 2002US-0358656P.  
 PR 21-FEB-2002; 2002US-0359122P.  
 PR 22-FEB-2002; 2002US-0358978P.  
 PR 22-FEB-2002; 2002US-0359034P.  
 PR 22-FEB-2002; 2002US-0359035P.  
 PR 27-FEB-2002; 2002US-0359121P.  
 PR 27-FEB-2002; 2002US-0359964P.  
 PR 01-MAR-2002; 2002US-0360858P.  
 PR 12-MAR-2002; 2002US-0363430P.  
 PR 12-MAR-2002; 2002US-0363676P.  
 PR 10-APR-2002; 2002US-0371346P.  
 PR 10-MAY-2002; 2002US-0379444P.

PR 04-JUN-2002; 2002US-00379444.  
 XX  
 PA (CUTRA-) CUTRAGEN CORP.  
 PI Agee ML, Anderson DW, Berghs C, Caeman SJ, Catterton E,  
 PI DiIppio VA, Edinger SR, Eissen A, Ellemann K, Gangolli EA,  
 PI Gerlach VL, Gorman L, Guo X, Herrmann JL, Hjalte T, Ji W, Kekuda R,  
 PI Khrantsov NV, Li L, Liu X, Malyanek UM, Miller CF, Millet I,  
 PI Ort T, Padigaru M, Patturajan M, Pena CFA, Rastelli L, Rieger DK,  
 PI Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G, Spaderna SK,  
 PI Spytek KA, Stone DJ, Vernet CAM, Zhong H, Zhong W, Alsdorok JP,  
 PI Burgess CE, Lepley DW;  
 XX  
 DR WPI; 2003-210149/20.  
 DR N-PSDB; ADC10203.  
 XX  
 PT New isolated NOVX polypeptides and nucleic acid molecules useful for  
 PT treating, preventing and diagnosing pathological conditions with NOVX-  
 PT associated disorders, such as cancer, obesity, diabetes and inflammatory  
 PT or CNS diseases.  
 PS Claim 1, SEQ ID NO 226; 772bp; English.  
 XX  
 CC The invention relates to novel isolated polypeptides, mature form of the  
 CC polypeptide, a sequence that is 95% identical to the polypeptide or the  
 CC polypeptide comprising one or more conservative substitutions. The NOVX  
 CC polypeptide is useful for treating or preventing a pathology associated  
 CC with the polypeptide e.g. disorders associated with aberrant expression  
 CC or activity of the polypeptide, such as cancer, diabetes, obesity, and  
 CC endocrine, CNS and inflammatory disorders. They can also be used in  
 CC various detection and screening assays, chromosome mapping, tissue typing  
 CC and predictive medicine. This sequence corresponds to one of the  
 CC polypeptides of the invention.  
 XX  
 SQ Sequence 272 AA;  
 Query Match 82.7%; Score 1378; DB 7; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-136;  
 Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 45 QESNLSIQALERSRDDILKRLYEKAAYDGLSKMIGPPDADLVNTNIIQADEPTTLTN 104  
 DB 5 QESNLSIQALERSRDDILKRLYEKAAYDGLSKMIGPPDADLVNTNIIQADEPTTLTN 64  
 QY 105 ADDLSVSGKDYGALKDVIYINANPASPPSLVLHRLCEHFRVLSYVHTSSVSVPEEN 164  
 DB 65 ADDLSVSGKDYGALKDVIYINANPASPPSLVLHRLCEHFRVLSYVHTSSVSVPEEN 124  
 QY 165 LKCFGEONKKQPRQDYQLGFTLLIKWNVKPTOMKFSIQTMCPTEGEGNIARFLFSLFGQK 224  
 DB 125 LKCFGEONKKQPRQDYQLGFTLLIKWNVKPTOMKFSIQTMCPTEGEGNIARFLFSLFGQK 184  
 QY 225 HNAVNAATLIDSWDIAIFOLKEGSSKEKAAYFRSMNSALGKSPWLAGNELTVADVLMVSV 284  
 DB 185 HNAVNAATLIDSWDIAIFOLKEGSSKEKAAYFRSMNSALGKSPWLAGNELTVADVLMVSV 244  
 QY 285 LQOIGGCSYTVPANVQRMRSCEENLAPF 312  
 DB 245 LQOIGGCSYTVPANVQRMRSCEENLAPF 272  
 RESULT 6  
 ADR86553  
 ID ADR86553 standard; protein; 229 AA.  
 AC ADR86553;  
 XX  
 DT 18-NOV-2004 (first entry)  
 DE 84-312 amino acid sequence of p38/JTV-1 protein.  
 KW p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.  
 XX

XX	Hom	sapiens.
XX	EP1454628-A2.	
XX	08-SEP-2004.	
XX	09-SEP-2003; 2003EP-00020344.	
XX	03-MAR-2003; 2003KR-00013058.	
XX	(UYSE-) UNIV SEOUL NAT IND FOUND.	
XX	Kim S, Park B;	
XX	WPI; 2004-627822/61.	
XX	N-PSDB; ADR86550.	
XX	New isolated p38/JTV-1 protein, useful as medicament for treating cancer and cutaneous or intraocular melanoma, as well as for screening new anticancer agents.	
XX	Claim 5; SEQ ID NO 6; 47pp; English.	
XX	The present invention relates to an isolated p38/JTV-1 protein for use as medicament. The p38/JTV-1 protein or the pharmaceutical composition is useful as medicament for treating breast cancer, large intestinal cancer, lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer, cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer, rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma, endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma, Hodgkin's disease, seopharyngeal cancer, small intestine cancer, endocrine cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue tumour, urethral cancer, penile cancer, prostate cancer, chronic or acute leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary adenoma, or their combination. The protein is useful as a target for screening new anticancer agents. The present sequence represents the 84-312 amino acid sequence of p38/JTV-1 protein.	
XX	Sequence 229 AA;	
XX	Query Match 71.4%; Score 1191; DB 8; Length 229;	
XX	Best Local Similarity 100.0%; Pred. No. 7.2e-117;	
XX	Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
XX	84 DADLVNTIIQADEPTTLTNALDINSVLGKDYGALKDIVIANPASPPLSLIVHRLIC 143	
XX	1 DADLVNTIIQADEPTTLTNALDINSVLGKDYGALKDIVIANPASPPLSLIVHRLIC 60	
XX	EHFRLSVTHTHSSSVKSVPENLKCGEGENKKOPRODYGFTLIKNVPTKQMSIOT 203	
XX	61 EHFRLSVTHTHSSSVKSVPENLKCGEGENKKOPRODYGFTLIKNVPTKQMSIOT 120	
XX	204 MCPLEGEGNIAFLFSLFGQKNAVNATLIDSMVDIAIQLEKGSSEKRAVFRSMNSAL 263	
XX	121 MCPLEGEGNIAFLFSLFGQKNAVNATLIDSMVDIAIQLEKGSSEKRAVFRSMNSAL 180	
XX	264 GKSPWLAGNELTVADVILMSVLQIQGGCVTVPANVRMRSCENLAPF 312	
XX	181 GKSPWLAGNELTVADVILMSVLQIQGGCVTVPANVRMRSCENLAPF 229	
XX	RESULT 7	
XX	ADR86552	
XX	ADR86552 standard; protein; 161 AA.	
XX	ADR86552;	
XX	18-NOV-2004 (first entry)	

Query Match	49.94; Score 832; DB 8; Length 161;
Best Local Similarity 100.0%; Pident. No. 3,9e-79;	
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MPMVQVPRYHGGGAPLAVELLPTCMYRPLPNVHGRSGYAPAGAHQOESNSTLOLESROD 60	
DB 1 MPMVQVPRYHGGGAPLAVELLPTCMYRPLPNVHGRSGYAPAGAHQOESNSTLOLESROD 60	
QY 61 DILKRLVELKAAVGLSKMIDTPADLDVTNIIOADEPTLTITNALDINSVLGKDYGALK 120	
DB 61 DILKRLVELKAAVGLSKMIDTPADLDVTNIIOADEPTLTITNALDINSVLGKDYGALK 120	
QY 121 DIVINANPASPPSLTLVLRLLCEHFRVLSTVTHSSVSKSV 161	
DB 121 DIVINANPASPPSLTLVLRLLCEHFRVLSTVTHSSVSKSV 161	
RESULT 8	
ABP01502	
ID ABP01502 standard; protein; 51 AA.	
AC ABP01502;	
XX	
DT 24-JUN-2002 (first entry)	
XX	

DE Human ORFX protein sequence SEQ ID NO:2986.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumor; hemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US010836.

PR 30-MAY-2000; 2000US-0206132P.

PR 29-AUG-2000; 2000US-0228716P.

PA (CURA-) CURAGEN CORP.

PI Shinketsu RA, Leach MD;

PI WPI, 2002-106308/14.

DR N-PSDB; ABLN17254.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
PS Disclosure; SEQ ID NO 2986; 1037pp; English.

XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-1191 (see Table 1  
CC in the specification). ABLN15762 to ABLN27252 encode the human ORFX  
CC proteins given in ABLN0010 to ABLN1500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 51 AA;

Query Match 14.8%; Score 247; DB 5; Length 51;  
Best Local Similarity 96.1%; Pred. No. 4.5e-16;  
Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 74 DGLSMIQTTPADLDVTNIIQADEPTLTITNALDINSVLGKDYGALKDVIY 124  
DB 1 DGLSMIQTTPADLDVTNIIQADEPTLTITNALDINSVLGKDYGALKDVIY 51

RESULT 9  
ABR62468

ID ABR62468 standard; protein; 334 AA.

XX ABR62468;

AC 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 14196.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI WPI, 2001-656860/75.

DR N-PSDB; ABL06571.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 14196; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR57737-  
CC ABR72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 334 AA;

Query Match 12.9%; Score 214.5; DB 4; Length 334;  
Best Local Similarity 25.6%; Pred. No. 2.8e-13;  
Matches 85; Conservative 50; Mismatches 120; Indels 77; Gaps 10;

QY 18 VELPTCMTRLPNV---HRSYGPAPGAGHVOESNLSLQA----- 54  
DB 13 IKLPFCMPPLKNVSLAADSLSGSSSTASTSASTSSCLANRIDRGRAATGALDLS 72  
QY 55 -----LESRODDILIKRYELKAAVDGSKMIGTPDADLDVTNIIQADEPT 99  
DB 73 LGRQYORLKDKDYSVAARQEVKQLEBLAAQGGQR-----AGLVCQ--KTFOHT 123  
QY 100 TTTNALDINSVLGKDYGALKDVIYNNPASPPLSLVLRLLCEHFRVLSVHTSSVYK 159  
DB 124 TAFQNG-----DLKEVPLQDVVINGHPNFIYALLALNARNAVITIVKTFHSTWA 176  
QY 160 SV-----PENLKCGEQNKQKQRPDYQGLFLTKWNPVKTKQMFSGTMCPIEGEGN 212  
DB 177 DIGPAAEFENAKV--PVNPALPK---ISVTLWKNCHEHTEMISSTMYVPYIGEVN 230  
QY 213 IARFLFSLFGQKHAANAVATL---IDSVWDIAIQLKEGSSKEKAAVFRSMNSALGSPML 269  
DB 231 IIRYLGRVGAPEAYRYEGSPCLNEIDVLVDICVLLKCNHKTQVANVRLDKRLQKQYF 290  
QY 270 AGNELTVADVVLVMSVLQIGGCSVTVPANVR 301



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 23, 2005, 13:36:20 ; Search time 17.81 Seconds  
(without alignments)  
1341.256 Million cell updates/sec

Title: US-10-622-817-3

Perfect score: 1667

Sequence: 1 MPWQVKKPHYGAGAPLRYEL.....RWRSCENLAPPTALKLK 320

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667	100.0	341	4	US-09-949-016-11312
2	1630	97.8	312	2	US-08-518-862C-2
3	103	6.2	1512	3	US-09-443-184-48
4	97	5.8	559	2	US-08-756-317-10
5	97	5.8	559	4	US-09-091-609-4
6	97	5.8	5215	3	US-09-105-537-2
7	95	5.7	559	4	US-09-821-016-1
8	95	5.7	559	4	US-10-259-632-1
9	95	5.7	559	4	US-10-266-787-1
10	92	5.5	2954	4	US-09-150-867-1
11	91	5.5	474	4	US-09-248-796A-20321
12	90.5	5.4	359	3	US-09-540-824-2
13	90.5	5.4	597	1	US-08-399-696-102
14	88.5	5.3	1007	4	US-09-538-092-736
15	88	5.3	427	4	US-09-949-016-11178
16	88	5.3	724	4	US-09-900-920-62
17	88	5.3	883	4	US-09-876-239-2
18	88	5.3	883	4	US-10-289-779B-2
19	88	5.3	914	4	US-09-976-239-4
20	88	5.3	914	4	US-10-289-779B-4
21	86.5	5.2	761	3	US-09-012-710-13
22	86.5	5.2	761	3	US-09-556-273-13
23	86	5.2	211	4	US-09-328-352-6834
24	86	5.2	351	4	US-08-178-257-6
25	86	5.2	471	3	US-09-134-001C-4904
26	86	5.2	1365	3	US-09-376-330-18
27	86	5.2	1365	6	5194600-4

28	86	5.2	1365	6	5194600-4	Patent No. 5194600
29	85.5	5.1	1440	3	US-09-357-251-37	Sequence 37, App1
30	85	5.1	559	4	US-10-253-509-1	Sequence 1, App1
31	85	5.1	1335	2	US-08-469-537A-97	Sequence 97, App1
32	84.5	5.1	579	4	US-09-107-433-448	Sequence 448, App1
33	84.5	5.1	829	4	US-07-670-611-2	Sequence 2, App1
34	84.5	5.1	829	1	US-08-220-674-2	Sequence 2, App1
35	84.5	5.1	829	1	US-08-445-186-2	Sequence 2, App1
36	84.5	5.1	829	1	US-08-446-549-2	Sequence 2, App1
37	84.5	5.1	829	2	US-08-446-550-2	Sequence 2, App1
38	83.5	5.0	778	4	US-09-583-110-3930	Sequence 3930, App1
39	83.5	5.0	786	4	US-09-107-433-3893	Sequence 3893, App1
40	83	5.0	559	2	US-09-052-339-1	Sequence 1, App1
41	83	5.0	559	3	US-09-385-742B-1	Sequence 1, App1
42	83	5.0	559	4	US-09-989-786-1	Sequence 1, App1
43	83	5.0	767	3	US-08-836-567-8	Sequence 8, App1
44	83	5.0	767	4	US-09-606-304-8	Sequence 8, App1
45	82.5	4.9	222	4	US-09-248-796A-18146	Sequence 18146, App1

#### ALIGNMENTS

```
RESULT 1
US-09-949-016-11312
; Sequence 11312, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11312
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11312

Query Match      100.0%; Score 1667; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 2.4e-181;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPWQVKKPHYGAGAPLRYELPTCMYRLPNVHGRSYGPAPGAGHVOEESNLISQALESSROD 60
        |||||||
DB      22 MPWQVKKPHYGAGAPLRYELPTCMYRLPNVHGRSYGPAPGAGHVOEESNLISQALESSROD 81

QY      61 DLTKLLEYLKAAYVDGSKMIOPTPDADLDVTNIIQADEPTTLTNALDLSVIGKDGALK 120
        |||||||
DB      82 DLTKLLEYLKAAYVDGSKMIOPTPDADLDVTNIIQADEPTTLTNALDLSVIGKDGALK 141

QY      121 DIVINANPASPPLSLVLRLLCEHFRVLTSTHTSSVSVSPENLLKCGEONKQKOPROD 180
        |||||||
DB      142 DIVINANPASPPLSLVLRLLCEHFRVLTSTHTSSVSVSPENLLKCGEONKQKOPROD 201

QY      181 YOLGFTLLMKVNPKTQMKFSIQTMCPRIEGEGNIARFLFSLFQKNAVNATLIDSVNDIA 240
        |||||||
DB      202 YOLGFTLLMKVNPKTQMKFSIQTMCPRIEGEGNIARFLFSLFQKNAVNATLIDSVNDIA 261

QY      241 IFOLGEGSKKKAAPFRSNNSALGKSPMLAGNELTVADVLVMSVLOQTGGCVTPVANNQ 300
        |||||||
DB      262 IFOLGEGSKKKAAPFRSNNSALGKSPMLAGNELTVADVLVMSVLOQTGGCVTPVANNQ 321

QY      301 RWRSCENLAPPTALKLK 320
```

Db 322 RMWSCENIAPFNTALTKLK 341

## RESULT 2

US-08-518-862C-2  
; Sequence 2, Application US/0851862C  
; Patent No. 5843757  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: Nicolaides, Nicholas C.  
; TITLE OF INVENTION: Human JTV1 Gene Overlaps PMS2 Gene  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/518, 862C  
; FILING DATE: 24-AUG-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 01107.49697  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9239  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 312 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-518-862C-2

Query Match 97.8%; Score 1630; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 3.5e-177;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPWYQVYKPHGGAPLRVEIPCTMYRLPVNHGSRVSGPAPAGHVOEESNLSTQALESRD 60  
1 MPWYQVYKPHGGAPLRVEIPCTMYRLPVNHGSRVSGPAPAGHVOEESNLSTQALESRD 60  
61 DIKRIYELKAAVDGSKMIOPTPDADLDYNTIIQADEPTTLTNNALDINSVIGKDYALK 120  
61 DIKRIYELKAAVDGSKMIOPTPDADLDYNTIIQADEPTTLTNNALDINSVIGKDYALK 120  
121 DIYINNPASPSLILVLRLLCEHFRVLSVTHSSVSVSEBENLKCFGEONKKOPROD 180  
121 DIYINNPASPSLILVLRLLCEHFRVLSVTHSSVSVSEBENLKCFGEONKKOPROD 180  
121 DIYINNPASPSLILVLRLLCEHFRVLSVTHSSVSVSEBENLKCFGEONKKOPROD 180  
181 YOLGFTLWKVNPQKQKFSIQTMCPRIEGEINAFPLSLFGQKHNANVATLIDSWDIA 240  
181 YOLGFTLWKVNPQKQKFSIQTMCPRIEGEINAFPLSLFGQKHNANVATLIDSWDIA 240  
181 YOLGFTLWKVNPQKQKFSIQTMCPRIEGEINAFPLSLFGQKHNANVATLIDSWDIA 240  
241 IFOLKEGSSKEKAQAVFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCSVTVPANQ 300  
241 IFOLKEGSSKEKAQAVFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCSVTVPANQ 300  
241 IFOLKEGSSKEKAQAVFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCSVTVPANQ 300  
301 RMWSCENIAPF 312  
301 RMWSCENIAPF 312

## RESULT 3

US-09-443-184-48  
; Sequence 48, Application US/09443184A  
; Patent No. 6372431  
; GENERAL INFORMATION:  
; APPLICANT: Cunningham, Mary Jane  
; APPLICANT: Zweigler, Gary  
; APPLICANT: Kaseer, Matthew R.  
; APPLICANT: Panzer, Scott  
; APPLICANT: Sellhammer, Jeffrey J.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Mariah  
; APPLICANT: Azimzai, Yalda  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS  
; FILE REFERENCE: PC-0007 US  
; CURRENT APPLICATION NUMBER: US/09/443,184A  
; NUMBER OF SEQ ID NOS: 138  
; SOFTWARE: PERL Program  
; SEQ ID NO 48  
; LENGTH: 1512  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1  
US-09-443-184-48

Query Match 6.2%; Score 103; DB 3; Length 1512;  
Best Local Similarity 22.1%; Pred. No. 0.12;  
Matches 45; Conservative 30; Mismatches 69; Indels 60; Gaps 9;

Db 124 INNPASPSLILVLRLLCEHFRVLSVTHSSVSVSEBENLKCFGEONKKOPRODYL 183  
6 LTVNSGDPPLGAL---LAVEHYKDVSI---SVEGKENILH----- 41  
Db 184 GFTLWKVNPQKQKFSIQTMCPRIEGEINAFPL-----FSLFGQKHNANVATLIDSWV 237  
42 ---VSENVIFTDV-----NSILRYLRVATVATGAGYS--NLMEHTETIDHWL 82  
Qy 238 DIAIFOLKEGSSKEKAQAVFRSMNSALGKSPWLAGNELTVADVLMVSLQOIGGCS----- 292  
83 EFSATKL--SSCDSFSTINELNHCLSLRTYLVGNSLSLADLCVWATLK--GNAAWQSO 138  
Qy 293 --VTVPANQRMWSCENIAPFNT 314  
Db 139 KKKKAPVHVWRMFGFLAQOAFOS 162

## RESULT 4

US-08-756-317-10  
; Sequence 10, Application US/08756317  
; Patent No. 5843694  
; GENERAL INFORMATION:  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Stark, David M.  
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:57:26 ; Search time 50.2443 Seconds  
(without alignments)  
2084.158 Million call updates/sec

Title: US-10-622-817-3  
Perfect score: 1667  
Sequence: 1 MPWYVKRPHYGAGAPLRYVEL.....RWRSCENLAPNTALXLLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1630	97.8	312	US-10-463-676-4	Sequence 4, Appl1
2	1191	71.4	229	US-10-463-676-6	Sequence 6, Appl1
3	832	49.9	161	US-10-463-676-5	Sequence 5, Appl1
4	102	6.1	201	US-10-369-493-11717	Sequence 13717, A
5	101	6.1	826	US-09-746-660A-92	Sequence 92, Appl1
6	101	6.1	833	US-09-746-660A-90	Sequence 90, Appl1
7	101	6.1	1221	US-09-919-891-2	Sequence 2, Appl1
8	101	6.1	1221	US-09-738-626-5163	Sequence 5163, Ap
9	101	6.1	1221	US-10-450-055-2	Sequence 2, Appl1
10	101	6.1	1221	US-10-494-675-156	Sequence 156, App
11	100.5	6.0	687	US-10-425-114-58897	Sequence 58897, A
12	100.5	6.0	1200	US-10-283-122A-53993	Sequence 53993, A
13	97.5	5.8	652	US-10-264-049-3001	Sequence 3001, Ap

14	97.5	5.8	711	US-10-437-963-175483	Sequence 175483,
15	97	5.8	559	US-09-364-847-21	Sequence 21, Appl1
16	97	5.8	856	US-09-364-847-33	Sequence 33, Appl1
17	97	5.8	856	US-09-364-847-35	Sequence 35, Appl1
18	97	5.8	5215	US-09-861-289-2	Sequence 2, Appl1
19	97	5.8	5215	US-09-860-846-2	Sequence 2, Appl1
20	97	5.8	5215	US-09-988-384B-2	Sequence 2, Appl1
21	97	5.8	5215	US-09-836-821-2	Sequence 2, Appl1
22	97	5.8	5215	US-10-271-889-45	Sequence 45, Appl1
23	96.5	5.8	914	US-10-437-963-153870	Sequence 153870,
24	96.5	5.8	1148	US-10-655-799-29	Sequence 29, Appl1
25	96	5.8	606	US-10-437-963-115189	Sequence 115189,
26	95.5	5.7	455	US-10-425-114-42633	Sequence 42633, A
27	95	5.7	559	US-09-821-016-1	Sequence 1, Appl1
28	95	5.7	559	US-09-820-952A-1	Sequence 1, Appl1
29	95	5.7	559	US-10-218-519-1	Sequence 1, Appl1
30	95	5.7	559	US-10-259-632-1	Sequence 1, Appl1
31	95	5.7	559	US-10-266-787-1	Sequence 1, Appl1
32	95	5.7	559	US-10-252-518-1	Sequence 1, Appl1
33	95	5.7	559	US-10-914-244-1	Sequence 1, Appl1
34	95	5.7	722	US-10-369-483-10013	Sequence 10013, A
35	95	5.7	936	US-10-032-585-7442	Sequence 7442, Ap
36	94.5	5.7	955	US-10-282-122A-69912	Sequence 69912, A
37	94	5.6	203	US-10-369-493-8604	Sequence 8604, Ap
38	94	5.6	468	US-10-424-599-260388	Sequence 260388,
39	94	5.6	1440	US-10-437-963-136277	Sequence 136277,
40	93.5	5.6	569	US-10-437-963-123904	Sequence 123904,
41	93.5	5.6	733	US-10-424-599-243766	Sequence 243766,
42	92	5.5	739	US-10-424-599-251522	Sequence 251522,
43	92	5.5	1206	US-10-282-122A-63838	Sequence 63838, A
44	91.5	5.5	1149	US-10-655-799-20	Sequence 30, Appl1
45	91.5	5.5	1192	US-10-282-122A-64644	Sequence 64644, A

## ALIGNMENTS

RESULT 1  
US-10-463-676-4  
; Sequence 4, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Park, Bum-Joon  
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; TITLE OF INVENTION: for Screening Pharmaceutical Composition for Treating Cancer  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; PRIOR FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FaastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(312)  
; OTHER INFORMATION: 1-312 amino acid sequence of p38/JTV-1  
US-10-463-676-4  
Query Match 97.8%; Score 1630, DB 16; Length 312;  
Best Local Similarity 100.0%; Pred. No. 6.4e-153;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPWYVKRPHYGAGAPLRYVELPTCMYRLPVHRSYGPAGAGHVEESNLSTQALSSRD 60  
DB 1 MPWYVKRPHYGAGAPLRYVELPTCMYRLPVHRSYGPAGAGHVEESNLSTQALSSRD 60  
QY 61 DILKRLYEKAVDGSIKMIGTPDDADLVNTIIQADEPTTLTNNALDNLVSGKDYGALK 120

Db 61 DILKRLVELKAAVDSLSKMIQTPEADLDVTNIIQADEPTTLTNNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPPLSLIVLHRLICEHFRVLSVTHSSVSVENILKCGEONKKQPRD 180  
Db 121 DIVINANPASPPLSLIVLHRLICEHFRVLSVTHSSVSVENILKCGEONKKQPRD 180  
QY 181 YOLGFTLWKVNPKTOMKFSIQTMCPIEBEGNIARLFSLFGQKNAVANATLIDSWDIA 240  
Db 181 YOLGFTLWKVNPKTOMKFSIQTMCPIEBEGNIARLFSLFGQKNAVANATLIDSWDIA 240  
QY 241 IFOLKEGSSKEKAARVRSNMSALGKSPMLAGNELTVADVILMSVLCQIGCSVTVPANQ 300  
Db 241 IFOLKEGSSKEKAARVRSNMSALGKSPMLAGNELTVADVILMSVLCQIGCSVTVPANQ 300  
QY 301 RWRSCENLAPF 312  
Db 301 RWRSCENLAPF 312

## RESULT 2

US-10-463-676-6  
; Sequence 6, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; APPLICANT: Park, Bum-Joon  
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; PRIOR FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(229)  
; OTHER INFORMATION: 84-312 amino acid sequence p38/JTV-1  
US-10-463-676-6

Query Match 71.4%; Score 1191; DB 16; Length 229;  
Best Local Similarity 100.0%; Pred. No. 1.5e-109; Indels 0; Gaps 0;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 DADLDVTNIIQADEPTTLTNNALDINSVLGKDYGALKDIVINANPASPPLSLIVLHRLIC 143  
Db 1 DADLDVTNIIQADEPTTLTNNALDINSVLGKDYGALKDIVINANPASPPLSLIVLHRLIC 60  
QY 144 EHFRLVSVTHSSVSVENILKCGEONKKQPRDYOGLTLLWKVNPKTOMKFSIQT 203  
Db 61 EHFRLVSVTHSSVSVENILKCGEONKKQPRDYOGLTLLWKVNPKTOMKFSIQT 120  
QY 204 MCPITGEGNIARLFSLFGQKNAVANATLIDSWDIAIFOLKEGSSKEKAARVRSNMSAL 263  
Db 121 MCPITGEGNIARLFSLFGQKNAVANATLIDSWDIAIFOLKEGSSKEKAARVRSNMSAL 180  
QY 264 GKSPMLAGNELTVADVILMSVLCQIGCSVTVPANQWRMRSCENLAPF 312  
Db 161 GKSPMLAGNELTVADVILMSVLCQIGCSVTVPANQWRMRSCENLAPF 229

## RESULT 3

US-10-463-676-5  
; Sequence 5, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; APPLICANT: Park, Bum-Joon

; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; PRIOR FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 161  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(161)  
; OTHER INFORMATION: 1-161 amino acid sequence of p38/JTV-1  
US-10-463-676-5

Query Match 49.9%; Score 832; DB 16; Length 161;  
Best Local Similarity 100.0%; Pred. No. 3.9e-74; Indels 0; Gaps 0;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMYQVKPYHGGAPLRYELPTCMYRLPNVHGRSYGAPAGAHVQESNLSIQALESRD 60  
Db 1 MEMYQVKPYHGGAPLRYELPTCMYRLPNVHGRSYGAPAGAHVQESNLSIQALESRD 60  
QY 61 DILKRLVELKAAVDSLSKMIQTPEADLDVTNIIQADEPTTLTNNALDINSVLGKDYGALK 120  
Db 61 DILKRLVELKAAVDSLSKMIQTPEADLDVTNIIQADEPTTLTNNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPPLSLIVLHRLICEHFRVLSVTHSSVSV 161  
Db 121 DIVINANPASPPLSLIVLHRLICEHFRVLSVTHSSVSV 161

## RESULT 4

US-10-369-493-13717  
; Sequence 13717, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 13717  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Pseudomonas fluorescens  
US-10-369-493-13717

Query Match 6.1%; Score 102; DB 15; Length 201;  
Best Local Similarity 30.5%; Pred. No. 0.17; Indels 22; Gaps 3;  
Matches 32; Conservative 15; Mismatches 36; Indels 22; Gaps 3;

QY 211 GNIRFLFSLFGQKNAVNA-TLIDSWDIAIFOLKEGSSKEKAARVRSNMSALGKSPWL 269  
Db 106 GPAARLITVGAFFNAEEVITRAHNFUKV-----MDLEIGKTPYL 146  
QY 270 AGNELTVADVILMSVLCQIGCSVTVPANQWRMRSCENLAPF 312  
Db 147 AGTEFTIDVSAVYIAHAPGNVSLDDYANVRAWLRIEALPGF 191

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:34:50 ; Search time 13.3213 Seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-3

Perfect score: 1667

Sequence: 1 MEMYGVKRYHGGAPLRLVEL.....RWRSCENLAPFTALKLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	7.5	719	2	T52043
2	111	6.7	728	2	T01200
3	105.5	6.3	222	2	G82441
4	97.5	5.8	218	2	A29036
5	97.5	5.8	645	2	JC5517
6	97	5.8	559	2	A3604
7	96.5	5.8	174	2	S74241
8	96	5.8	1675	2	S70770
9	94	5.6	218	2	B29231
10	93	5.6	393	2	D64391
11	92	5.5	210	2	B85594
12	92	5.5	210	2	P64821
13	92	5.5	210	2	P90743
14	92	5.5	744	2	A82703
15	92	5.5	764	2	E97485
16	92	5.5	867	2	S72842
17	92	5.5	1036	2	AF2338
18	92	5.5	1306	2	E87072
19	92	5.5	2954	2	T14156
20	91.5	5.5	218	1	XURIG4
21	91.5	5.5	1192	2	G70513
22	91	5.5	2279	2	T42531
23	91	5.5	2280	2	T38906
24	90.5	5.4	359	2	T37921
25	90	5.4	710	2	S67098
26	89.5	5.4	218	2	B34159
27	89.5	5.4	1365	1	BVBK5
28	89	5.3	2480	2	D84904
29	88.5	5.3	209	1	XUFP11

30	88.5	5.3	1007	2	S48535	rho-type GTPase-ac
31	88	5.3	207	2	D72272	conserved hypothe
32	88	5.3	218	2	B28946	glutathione transf
33	88	5.3	437	2	I51238	translation elonga
34	87.5	5.2	214	2	B46881	glutathione transf
35	87.5	5.2	1065	2	A43421	SEC8 protein - yea
36	87.5	5.2	1292	2	T31462	probable magnesium
37	87.5	5.2	1714	1	S18644	multifunctional am
38	87.5	5.2	2586	2	T21676	hypotheical prote
39	87	5.2	586	2	C71695	ctp synthase (pyr
40	86.5	5.2	208	1	S43851	glutathione trans
41	86.5	5.2	379	2	D97198	methyl-accepting c
42	86.5	5.2	436	2	I51237	translation elonga
43	86.5	5.2	439	2	D86440	unknown protein [1
44	86.5	5.2	749	2	A57276	MAU1 protein - yea
45	86	5.2	377	2	P69172	hypotheical prote

## ALIGNMENTS

```

RESULT 1
T52043
probable glutamate-tRNA ligase (EC 6.1.1.17) [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #ext_change 09-Jul-2004
C/Accession: T52043
R/Day, I.S.; Golovkin, M.; Reddy, A.S.
Biochim. Biophys. Acta 1399, 219-24, 1998
A/Title: Cloning of the cDNA for glutamyl-tRNA synthetase from Arabidopsis thaliana.
A/Reference number: Z24836; MUID:9765600; PMID:9765600
A/Accession: T52043
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-719 <DAY>
A/Cross-references: UNIPROT:O82462; EMBL:AF067773; PIDD:AAC36469.1
A/Suprafamily: Yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology
C/Keywords: ligase

Query Match          7.5%; Score 125.5; DB 2; Length 719;
Best Local Similarity 27.3%; Pred. No. 0.017;
Matches 48; Conservative 26; Mismatches 53; Indels 49; Gaps 9;

QY 128 PASPPSLVLRRLCEHFRVLSTVTHSSVSPENLTKCFGEQNKQPRODYGFTL 187
DB 10 PSSPPSLVVALSLASPV-----TIDSSMAATTVSFFV---PSDGRKLN-----GATV 55

QY 188 IKKNVPTQMKPFSIQTMCPIDEGNIAFLFSLFGQKHNAVNATLIDSWVDIA-IFOLKE 246
DB 56 LIRVY-----GRSAKKLPPFYG--NNAPSSQIDENVYASVF--SS 93

QY 247 GSSKEKAIVFRSMNSALGKSPWLAGNELTVADVVLMSVLQIGGCVTVPAVQWR 302
DB 94 GSEFENAC--GRVDKYLESSFLVGHSHSIADVAIWSALAGTG-----QRW 137

RESULT 2
T01200
probable glutamate-tRNA ligase (EC 6.1.1.17) F21E10.12 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #ext_change 09-Jul-2004
C/Accession: T01200
R/Davieson, S.; Rohlfing, T.; David, M.; O'Brian, D.
submitted to the EMBL Data Library, April 1998
A/Description: The sequence of A. thaliana F21E10.
A/Reference number: Z14258
A/Accession: T01200
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-728 <DAY>
A/Cross-references: UNIPROT:O65253; EMBL:AF058914; NID:G3047074; PID:G3047084; GSPDB:GNO
A/Experimental source: cultivar Columbia
C/Genetics:

```

A:Gene: ATSP:F21E10.12

A:Map position: 5

A:Introns: 47/2: 89/3; 141/1; 503/3; 659/3

C:Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology

C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

F:223-499/Domain: glutamine-tRNA ligase homology <EGF>

Query Match 6.7%; Score 111; DB 2; Length 728;

Best Local Similarity 25.9%; Pred. No. 0.3;

Matches 48; Conservative 26; Mismatches 53; Indels 58; Gaps 10;

QY 128 PASPPISLVLRLLCEHFRVLSTVHTSSVKSVPENILKCGEQNKQRPQDQLGFTL 187

Db 10 PESPPLSVIALSLASPY---TIDSSAAATTVPSPV---FSDGRKAN-----GATV 55

QY 188 IKWNPVKQMKFSIQTMCPISBEGNIARFLPSFGQKNANVAT-----LIDSWD 238

Db 56 LRRV-----GRSARKLPDPFG--NNAPDSQVSTILCINMKIDRW 95

QY 239 IAFOLKESKSKRAAVFRSMNSALGKSPWLAGNELTVADVLMVLIQIGGSVTVPA 297

Db 96 YASVF--SSGSFEVAC--GRVDKLTBSTSTLVGHSLIADVAIWSALAGTG----- 143

QY 298 NVQRW 302

Db 144 --QRW 146

RESULT 3

G82441

probable glutathione S-transferase VCA0584 [imported] - vibrio cholerae (strain N16961)

C:Species: vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C:Accession: G82441

R:Heidelberg, J.F.; Bisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:2040683; PMID:10952301

A:Accession: G82441

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-222 <HEI>

A:Cross-references: UNIPROT:Q9NM05; GB:AE004389; GB:AE003853; NID:99657989; PIDD:AAF9648

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Gene: VCA0584

A:Map position: 2

C:Superfamily: hypothetical protein b2302

Query Match 6.3%; Score 105.5; DB 2; Length 222;

Best Local Similarity 25.7%; Pred. No. 0.17;

Matches 28; Conservative 22; Mismatches 38; Indels 21; Gaps 4;

QY 197 MKFSIQTMCPISBEGNIARFLPSFGQKNANVATLIDSWDIALFOLKESKSKRAAV 256

Db 97 LMFQMSGVPMQGANV---FYRPEKIQP---IDY-----QKEGRRLF 137

QY 257 RSMNSALGKSPWLAGNELTVADVLMVLIQIGGSVTVPAVQRW 303

Db 138 EVMDDQLAQNPLYAGDEYTIADTFPVRRIHMGISIDGLTHLQRW 186

RESULT 4

A29036

glutathione transferase (EC 2.5.1.18) Yb3 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 09-Jul-2004

C:Accession: A29036

R:Abrahamovitz, M.; Lisowsky, I.

J. Biol. Chem. 262, 7770-7773, 1987

A:Title: Selective expression of a unique glutathione S-transferase Yb3 gene in rat brai

A:Reference number: A29036; MUID:87222405; PMID:3584141

A:Accession: A29036

A:Molecule type: mRNA

A:Residues: 1-218 <ABR>

A:Cross-references: UNIPROT:P08009; GB:J02744; NID:G204512; PIDD:AAA1292.1; PID:G20451

C:Superfamily: glutathione transferase

C:Keywords: transferase

Query Match 5.8%; Score 97.5; DB 2; Length 218;

Best Local Similarity 25.6%; Pred. No. 0.78;

Matches 33; Conservative 24; Mismatches 41; Indels 31; Gaps 5;

QY 179 QDYOLGFTLIWKNVPTQMKFSIQTMCPISBEGNIARFLPSFGQKNANVAT----- 231

Db 49 EKFKLG--LDFPNLP-----YLDGSHKITQSNALRLT-----GRKNLCETEBERRRV 97

QY 232 -----LIDSWDIA-----IFOLKESKSKRAAVFRSMNSALGKSPWLAGNELTVAD 278

Db 98 DILENOLDNRNVLARLCYNDPFEKLPKGYLEOLPGMWRLYSEFLGKRPWPAQDKITFVD 157

QY 279 VVLMGVLQ 287

Db 158 FIAYDVLER 166

RESULT 5

JC5517

Gu/RNA helicase II binding protein - human

C:Species: Homo sapiens (man)

C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 05-Nov-1999

C:Accession: JC5517

R:Valdez, B.C.; Henning, D.; Perlaky, L.; Busch, R.K.; Busch, H.

Biochem. Biophys. Res. Commun. 234, 335-340, 1997

A:Title: Cloning and characterization of Gu/RN-II binding protein.

A:Reference number: JC5517; MUID:97320420; PMID:9177271

A:Accession: JC5517

A:Molecule type: mRNA

A:Residues: 1-645 <VAL>

A:Cross-references: GB:U78524; NID:91696006; PIDD:AA858488.1; PID:91696007

A>Note: It is uncertain whether Met-4 or Met-6 is the initiator

C:Comment: This protein is localized to the nucleus and interacts with Gu/RNA helicase I

C:Keywords: phosphoprotein

F:50-58/Region: nuclear location signal

F:362-374/Region: nuclear location signal

F:514-517, 551-554, 592-595, 606-609/Region: 4 residue repeats (N-T-S-L)

F:717,732,412,432,460,461,462,470,614,627/Binding site: phosphate (Ser) (covalent) #st

F:65,394,624/Binding site: phosphate (Thr) (covalent) #st

Query Match 5.8%; Score 97.5; DB 2; Length 645;

Best Local Similarity 20.5%; Pred. No. 3.7; Length 121; Gaps 20;

Matches 79; Conservative 56; Mismatches 129; Indels 121; Gaps 20;

QY 16 LRV-ELPTCK-YRLPVHGRSYGAPAGHVOESNTSLQLESRODDILKRLVELKAAV 73

Db 8 LRVESEQLVLLGAVGRNKRGRKHELLTKALHLKAG--CSFAVQMKIKELYYRRRPQK--- 62

QY 74 DGLSKMIQTPDDADVTNIIQADEPTTLTNMLDINSVIGKGYLKOIVINANPASP 133

Db 63 -----IMRP-ADLSIPNHSSPMATLSPS-----TIPQLTYDGHFASPL 102

QY 134 ---SLI-VLRLLCEHF-RVLSVTHSSVKSVP-----ENILK--CRGEQNKQO----- 176

Db 103 LPVSLILGPGHELEPLHTSALHPVAPDKLQGLRPYDLDLDELIKRTSLASDNRFRFTC 162

QY 177 -----PRDYOLGFTLIWKNVPTQMKFSIQ-----TMCPIEGEG----- 211

Db 163 FAPALTPQOVQOISSM--DISGTCDFTVQVQLFCLSETSCPOEDHFPFNNLVKVT 219

QY 212 -----NIARFLPSFGQKNANVATLIDSW-----DI 239

Db 220 KPCSLPGVLPPTKNGVEKRRSRPNTISLV-----RISTTVPTVITWSVTAELGRNYSW 274

QY 240 AIFOLKESG-----KEKAVFRSMNSALGKSPWLAGNELTVADVLMVLIQIGGC 291

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:33:04 ; Search time 60.8145 Seconds  
(without alignments)  
2694.512 Million cell updates/sec

Title: US-10-622-817-3

Perfect score: 1667

Sequence: 1 MPWGVKPYHGGCAPLRLV.....RMWRSCENLAPNTALKLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667	100.0	320	1	MCA2_HUMAN
2	1663	99.8	320	2	O96CZ5
3	1464	87.8	320	2	O8R010
4	1464	87.8	320	2	O8R2Y6
5	1444	86.6	320	1	MCA2_CRIGR
6	1260	75.6	280	2	O8R3V2
7	1144.5	68.7	311	2	O6DK86
8	1139.5	68.4	311	2	O6IN04
9	1132.5	67.9	311	2	O7Z1D7
10	887.5	53.2	321	2	O7T3C0
11	267.5	16.0	340	2	O7Q7A3
12	224	13.4	301	2	O7KUM5
13	224	13.4	313	2	O8T060
14	224	13.4	322	2	O6NMK4
15	214.5	12.9	334	1	MCA2_DROME
16	125.5	7.5	719	2	O82462
17	111	6.7	728	2	O65253
18	106.5	6.4	880	2	O8ERT1
19	106.5	6.4	913	2	O6DRB3
20	106.5	6.4	913	2	O6PRQ4
21	105.5	6.3	222	2	O9KM05
22	105	6.3	240	2	O8BTR1
23	105	6.3	1512	1	SYEP_MOUSE
24	105	6.3	1800	2	O7PRA2
25	104.5	6.3	922	2	O6TEM5
26	104	6.2	661	2	O9VB85
27	104	6.2	702	2	O8IHC6
28	104	6.2	841	2	O8IMQ2
29	103	6.2	328	2	O86X73
30	103	6.2	869	2	O6PD57
31	103	6.2	1213	2	O8FRD2

32	102.5	6.1	372	2	O87XN7	O87XN7 pseudomonas
33	102.5	6.1	715	2	O9LIZ8	O9LIZ8 oryza sativ
34	102	6.1	559	2	O8R023	O8R023 pseudomonas
35	102	6.1	559	2	O8RD25	O8RD25 pseudomonas
36	102	6.1	1486	2	O6TXE9	O6TXE9 rattus norv
37	101	6.1	1196	2	O6W580	O6W580 corynebacte
38	101	6.1	1221	2	O8NQD1	O8NQD1 corynebacte
39	100.5	6.0	454	2	O6UDR4	O6UDR4 mus musculu
40	100.5	6.0	1201	2	O6NH83	O6NH83 corynebacte
41	99.5	6.0	651	1	PIA1_MOUSE	O88907 mus musculu
42	99	5.9	203	2	O6W112	O6W112 thizobium s
43	98.5	5.9	222	2	O8EGM4	O8EGM4 shewanella
44	98.5	5.9	582	2	O6ECB1	O6ECB1 yersinia ps
45	97.5	5.8	217	1	GTW3_RAT	P08009 rattus norv

## ALIGNMENTS

RESULT 1  
MCA2\_HUMAN  
ID MCA2\_HUMAN STANDARD; PRT; 320 AA.  
AC Q13155; Q9P1L2;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Multisynthetase complex auxilliary component p38 (JTV-1 protein)  
DE (PRO092).  
GN Name=JTV1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96115582; PubMed=8666379;  
RA Nicolaides N.C., Kinzler K.W., Vogelstein B.;  
RT "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts  
and a novel overlapping gene";  
RL Genomics 29:329-334 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshuler S.F., Zeeberg B., Buelow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hultk F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.U., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP SEQUENCE OF 197-320 FROM N.A.  
RC TISSUE=Fetal liver;  
RX Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
RA Liu M., He F.;  
RT "Functional prediction of the coding sequences of 121 new genes  
deduced by analysis of cDNA clones from human fetal liver";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]

RP INTERACTION WITH FUBP1.  
RX MEDLINE=22716800; PubMed=12819782; DOI=10.1038/ng1182;  
RA Kim M.J., Park B.-J., Kang Y.S., Kim H.J., Park J.-H., Kang J.W.,  
RT Lee S.W., Han J.M., Lee H.-W., Kim S.,  
RT "Downregulation of FUS- binding protein and c-myc by tRNA synthetase  
RT cofactor p38 is required for lung cell differentiation.";  
RL Nat. Genet. 34:330-336(2003).  
CC -1- FUNCTION: Probable core protein of the multisynthetase complex  
CC that serves as a template for the assembly of the supramolecular  
CC structure. Mediates ubiquitination of FUBP1 and its degradation by  
CC the proteasome.  
CC  
CC -1- SUBUNIT: Component of the multisynthetase complex which is  
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the  
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary  
CC proteins, p18, p48 and p43. Binds FUBP1.  
CC  
CC -1- SIMILARITY: Contains 1 GST-like domain.  
CC  
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 312.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL, U24169; AAC50391.1; ALT FRAME.  
DR EMBL, BC028853; AAH02853.1; -  
DR EMBL, BC010156; AAH010156.1; -  
DR EMBL, AF11615; AAF71039.1; -  
DR H-InvDB, HIX0006460; -  
DR MIM, 600859; -  
DR InterPro: IPR010987; GST\_C-like.  
DR InterPro: IPR004046; GST\_Cterm.  
DR Pfam, PF00043; GST\_C; 1.  
DR Protein biosynthesis.  
SQ SEQUENCE 320 AA; 35349 MW; F253726B63C12B8B CRC64;  
Query Match 100.0%; Score 1667; DB 1; Length 320;  
Best Local Similarity 100.0%; Pred. No. 66-127;  
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPYQVPRYHGGAPLRVELPTCMYRLPNVHGRSGYPAGAGHVOESNLSLOAESROD 60  
DB 1 MPYQVPRYHGGAPLRVELPTCMYRLPNVHGRSGYPAGAGHVOESNLSLOAESROD 60  
QY 61 DILKRLYEELKAAVDGSLKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
DB 61 DILKRLYEELKAAVDGSLKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPLSLVLRHLCEHFRVLSVHTHSSVSKSVPENLLKCFEBOKKOPROD 180  
DB 121 DIVINANPASPLSLVLRHLCEHFRVLSVHTHSSVSKSVPENLLKCFEBOKKOPROD 180  
QY 181 YOLGFTLLIKWNPVPTQMKFSIQTCPIEGEGNIARFLFSFGOKHNAVNTLIDSWDIA 240  
DB 181 YOLGFTLLIKWNPVPTQMKFSIQTCPIEGEGNIARFLFSFGOKHNAVNTLIDSWDIA 240  
QY 241 IFOLKEGSSKEKAIVFRSNMSALGKSPMLAGNELTVADVILMSVLOOIGGCSVTVPANQ 300  
DB 241 IFOLKEGSSKEKAIVFRSNMSALGKSPMLAGNELTVADVILMSVLOOIGGCSVTVPANQ 300  
QY 301 RWRMSCENLAPNTALKLK 320  
DB 301 RWRMSCENLAPNTALKLK 320  
RESULT 2  
ID 096C25  
AC 096C25; 096C25; PRELIMINARY; PRT; 320 AA.

DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE U7V1.  
GN Name=U7V1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshlyuk S., Carninci P., Prange C.,  
RA Bosak S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Richards S., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.V., Huiyk S.W.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzyzanski M.I., Skalska U., Smalins D.E., Schmech A., Schen J.E.,  
RT Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, BC013630; AAH13630.1; -  
DR InterPro: IPR004046; GST\_Cterm.  
DR InterPro: IPR010987; GST\_C-like.  
DR Pfam, PF00043; GST\_C; 1.  
SQ SEQUENCE 320 AA; 35335 MW; 19F14BF58612E08 CRC64;  
Query Match 99.8%; Score 1663; DB 2; Length 320;  
Best Local Similarity 99.7%; Pred. No. 1,36-126;  
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MPYQVPRYHGGAPLRVELPTCMYRLPNVHGRSGYPAGAGHVOESNLSLOAESROD 60  
DB 1 MPYQVPRYHGGAPLRVELPTCMYRLPNVHGRSGYPAGAGHVOESNLSLOAESROD 60  
QY 61 DILKRLYEELKAAVDGSLKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
DB 61 DILKRLYEELKAAVDGSLKMIQTDPADLDVTNIIQADEPTTLTTNALDINSVLGKDYGALK 120  
QY 121 DIVINANPASPLSLVLRHLCEHFRVLSVHTHSSVSKSVPENLLKCFEBOKKOPROD 180  
DB 121 DIVINANPASPLSLVLRHLCEHFRVLSVHTHSSVSKSVPENLLKCFEBOKKOPROD 180  
QY 181 YOLGFTLLIKWNPVPTQMKFSIQTCPIEGEGNIARFLFSFGOKHNAVNTLIDSWDIA 240  
DB 181 YOLGFTLLIKWNPVPTQMKFSIQTCPIEGEGNIARFLFSFGOKHNAVNTLIDSWDIA 240  
QY 241 IFOLKEGSSKEKAIVFRSNMSALGKSPMLAGNELTVADVILMSVLOOIGGCSVTVPANQ 300  
DB 241 IFOLKEGSSKEKAIVFRSNMSALGKSPMLAGNELTVADVILMSVLOOIGGCSVTVPANQ 300  
QY 301 RWRMSCENLAPNTALKLK 320  
DB 301 RWRMSCENLAPNTALKLK 320



```

RESULT 3
OBR010 PRELIMINARY: PRT: 320 AA.
ID OBR010
AC OBR010;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Jtvl-pending protein.
GN Name=Jtvl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; and FVB/N-3; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor; C3;
RA Strausberg R.;
RN Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Mammary tumor;
RA Strausberg R.;
RN Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024410; AAH24410.1; -.
DR EMBL; BC026972; AAH26972.1; -.
DR MGD; MGI:2385237; Jtvl.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
SQ SEQUENCE 320 AA; 35396 MW; 1745D7EB4BC367D0 CRC64;

Query Match 87.8%; Score 1464; DB 2; Length 320;
Best Local Similarity 87.2%; Pred. No. 1.8e-110; Indels 0; Gaps 0;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

OY 1 MEMYGVKPYHGAGAPLRLVELPTCMYRLPNVHGSRSGYPAGAGHVEESNLSIQALESROD 60
DB 1 MEMYGVKPYHGAGAPLRLVELPTCMYRLPNVHGSRSGYPAGAGHVEESNLSIQALESROD 60
OY 61 DILKRLYLKKAADVGLSKMIGTPDADLDVTNIIQADEPTTLTNALDLSVYGLKDYGALK 120
DB 61 DILKRLYLKKAADVGLSKMIGTPDADLDVTNIIQADEPTTLTNALDLSVYGLKDYGALK 120
OY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLJKCGEQNKQKQPROD 180
DB 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLJKCGEQNKQKQPROD 180
OY 181 YOLGFTLWKVNPKTQMKFPIQTMCPIEEGENIARLFLFGQKHNAVNATLIDSVWDIA 240
DB 181 YOLGFTLWKVNPKTQMKFPIQTMCPIEEGENIARLFLFGQKHNAVNATLIDSVWDIA 240

```

```

DB 181 YOLGFTLWKVNPKTQMKFPIQTMCPIEEGENIARLFLFGQKHNAVNATLIDSVWDIA 240
OY 241 IOLKEGSKERKAAFRSNNSALGKSPMLAGNELTADVLVSVLQOIGCGSVTPVANYO 300
DB 241 MFOLEBSSKERKAAFRSNNSALGKSPMLVGNELTADVLVSVLQOIGCGSVTPVANYO 300
OY 301 RMRSCENLAPPNTAKLK 320
DB 301 RMRSCENLAPPNTAKLK 320
OY 301 RMRSCENLAPPNTAKLK 320
DB 301 RMRSCENLAPPNTAKLK 320

RESULT 4
OBR2Y6 PRELIMINARY: PRT: 320 AA.
ID OBR2Y6
AC OBR2Y6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Jtvl-pending protein.
GN Name=Jtvl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RA Strausberg R.;
RN Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026958; AAH26958.1; -.
DR MGD; MGI:2385237; Jtvl.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
SQ SEQUENCE 320 AA; 35423 MW; 1C21F1A74C9882B4 CRC64;

Query Match 87.8%; Score 1464; DB 2; Length 320;
Best Local Similarity 87.2%; Pred. No. 1.8e-110; Indels 0; Gaps 0;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

OY 1 MEMYGVKPYHGAGAPLRLVELPTCMYRLPNVHGSRSGYPAGAGHVEESNLSIQALESROD 60
DB 1 MEMYGVKPYHGAGAPLRLVELPTCMYRLPNVHGSRSGYPAGAGHVEESNLSIQALESROD 60
OY 61 DILKRLYLKKAADVGLSKMIGTPDADLDVTNIIQADEPTTLTNALDLSVYGLKDYGALK 120
DB 61 DILKRLYLKKAADVGLSKMIGTPDADLDVTNIIQADEPTTLTNALDLSVYGLKDYGALK 120
OY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLJKCGEQNKQKQPROD 180
DB 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSSVKSPENLJKCGEQNKQKQPROD 180

```

D <sub>b</sub>	121	DI V I N A N P A P P E S L I V H L I C E R Y R V S I V H T H S S Y K N V E N I V K F O G O A R K S R H E	180
Q <sub>y</sub>	181	Y O L G F T I I M N O V P K T O K E S I O T M C P I B E G N I A R F L E S L F G O K H N A V N A T L I D S W D I A	240
D <sub>b</sub>	181	Y O L G F T I I M N O V P K T O K E S V O T M C P I B E G N I A R F L E S L F G O K H N A V N A T L I D S W D I A	240
Q <sub>y</sub>	241	I F O L K E G S S K E K A A V R S M N S A L G S P M I A G E L T V A V L M S Y L O G I G G S V T V P A N V O	300
D <sub>b</sub>	241	M F O L R E G S S K E K A A V R S M N S A L G S P M I V G E L T V A V L M S Y L O G I G G S S G A A P T N V O	300
Q <sub>y</sub>	301	R M R S C E N A P P T A L K L K	320
D <sub>b</sub>	301	R M L K S C E N A P P S T A L O L K	320

RESULT 5	
MCA2_CRIGR	
ID MCA2_CRIGR	STANDARD;
AC 09WMV7.	PRT; 320 AA

DT 16-OCT-2001 (Rel. 4.0, Created)  
DT 16-OCT-2001 (Rel. 4.0, Last sequence update)  
DT 25-OCT-2004 (Rel. 4.5, Last annotation update)  
DE Multisynthetase complex auxiliary component p38.  
OS Cricetus griseus (Chinese hamster).  
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eukaryota; Euteheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetulus.  
OC NCBT\_TaxID=10029;  
OX [1]

RP SEQUENCE FROM N.A.  
RC TISSUE-Ovary;  
RA MEDLINE=99096915; PubMed=9878398; DOI=10.1006/jmbi.1998.2316;  
RX Quevillon S., Robinson J.-C., Berthomieu E., Slatecka M., Mizande M.;  
RT "Macromolecular assembly of aminoacyl-tRNA synthetases:  
RT identification of protein-protein interactions and characterization of  
RT a core protein.";  
RL J. Mol. Biol. 285:183-195(1999).  
CC -i- FUNCTION: probable core protein of the aminoacyl-tRNA synthetase  
CC complex.

that serves as a template for the assembly of the supramolecular structure.

-1-SUBUNIT: Component of the multisynthetase complex which is comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl, arginyl, and aspartyl-tRNA synthetases as well as three auxiliary proteins, p18, p48 and p43.

-1-SIMILARITY: Contains 1 GST-like domain.

-----  
This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its non-profit use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL: AF072727; AAD38422.1; -.  
InterPro: IPR010987; GST\_C like.  
InterPro: IPR004046; GST\_Cterm.  
Pfam: PF00043; GST\_C; 1.  
Protein biosynthesis.  
SEQUENCE 320 AA; 35433 MW; 6D24E033ABEC910A CRC64;  
O

Query Match	86.6%;	Score 1444;	DB 1;	Length 320;
Best Local Similarity	86.6%;	Pred. No. 7.6e-109;		
Matches 277;	Conservative 16;	Mismatches 37;		

0;	Gaps	0;	models	2/;	addmacchines	20;	.....
----	------	----	--------	-----	--------------	-----	-------

```

1 MPMYQVKPHHGGGAPRLVELPMYRLPNVHGRSGPAPGAGHVQESNLSLOLESROD 60
|||||
1 MPMYQVKSHHGGGAPRLVELPMYRLPNVHSGKTPSSDATTACVQTEPENNOCATITROD 60
|||||

```

61 DILKELYELKAVDGSKMIQTPDADIVTNIIOABEPTTLTNALDLSVIGKDYALK 120

Db	61	DIKRLKYLKAVDGLSMGHTPPADLDVTNIIQADEPVLTTNALDNLVLGKDYALK	120
Qy	121	DIVINANPASPPLSLVLRLLCEHFPVLSTVHTSHSVKSPENILKCFGEONKKOPROD	160
Db	121	DIVINANPASPPLSLVLRLLCEHFPVLSTVHTSHSVKSPENILKCFGEONKKOPROD	180
Qy	181	YOLGFTLLMKRVPTOMKKEFSGTMCPLFEGSGNINARFLFSLFGQGHANVNTLLIDSWDIA	240
Db	181	YOLGFTLLMKRVPTOMKKEFSGTMCPLFEGSGNINARFLFSLFGQGHANVNTLLIDSWDIA	240
Qy	241	IFOLKEGSSKEKAAPFRSMNSALGKSPWLTGNELTVADVLMGVLQOIGGCSYTPVANO	300
Db	241	MFOLKEGSSKEKAAPFRSMNSALGKSPWLTGNELTVADVLMGVLQOIGGCSYTPVANO	300
Qy	301	RANRSGCNLAPNTALIKL 320	
Db	301	RMLKSCENLVFPSTALOGLK 320	

RESULT 6	
Q8R3V2	
ID Q8R3V2	PRELIMINARY;
AC Q8R3V2.	PRT; 280 AA

DT 01-JUN-2002 (TEMBLrel, 21, Created)  
 DT 01-JUN-2002 (TEMBLrel, 21, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel, 25, Last annotation update)  
 DE Jvarkit.  
 GN Name=Jvarkit;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090.

RD SEQUENCE FROM N.A.  
RC STRAIN=EV3/N; TISSUE=Mammary tumor. C3;  
RX MEDLINE=23288557; PubMed=12477933; DOI=10.1073/pnas.242603899;  
RA Klausberg R. L., Feingold E.A., Grouse L.H., Deize J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.R., Schaefer C.P., Bhatt N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marasina K., Farmer A.C., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Cassanova T., Hong L.,

RA Brownstein M.J., Uselin T.B., Toshiyuki S., Carminci P., Franze C.,  
RA Raha S.S., Lofuellarlo N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bossek S.A., McGowan P.T., McKernan K.J., Malek U.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultx S.W.,  
RA Villalon D.K., Muzny D.M., Sodegberg E.J., Lu X., Gibbs R.A.,  
RA Falley J., Helton E., Kettelman M., Madan A.C., Shevchenko V., Bouffard G.G.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz U., Myers R.M., Butterfield V.S.,  
RA Krywinski M.J., Skalkida H., Schmutz U., Myers R.M., Butterfield V.S.,

RA Jones S.J., Mairia M.A., Shattell D.E., Schmetz A., Schein J.E.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
KL and mouse cDNA sequences. ",  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [21]

RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
R EMBL: BC024480.2,24480.1

DR MGD, MGI:2385237, Jv1.  
DR InterPro, IPR004046, GST\_Cterm.  
DR InterPro, IPR01987, GST\_C\_like.  
Pfam, PF00043, Gcm\_C\_1

SEQUENCE 280 AA; 31097 MW; E5BE81498983FD2E CRC64;

Query Match	75.64;	Score 1260;	DB 2;	Length 280;
Best Local Similarity	88.08;			

45 OEESNLSIQALESHQDDILKRLVELKAAVDSKMTQTEDADDVNTIIQADEPTLTIN 104  
Matches 243; Conservative 14; Mismatches 19; Indels 0; Gaps 0

Db 5 GQTSRPSLQALBSRQDDILKRLYELKAAVDGSKMHTPDADVTNITLQADEPTLATN 64  
Qy 105 ADDINSVLCGVKALDQYIYNNPASPPLSLVHLHLCBPRVSLVTHHSSVKSPEN 164  
Db 65 TLDLNSVLCGVKALDQYIYNNPASPPLSLVHLHLCBPRVSLVTHHSSVKSPEN 124  
Qy 165 LKCFGEQKKQKQPRODQYQGLTLMKQVPTQMKFSIQTMCPLEGSGNIAFLFSLFGOK 224  
Db 125 LKCFGEQKKQKQPRODQYQGLTLMKQVPTQMKFSIQTMCPLEGSGNIAFLFSLFGOK 184  
Qy 225 HNAVNATLIDSVWDIAIFQLKEGSSKEKAAVFRSNMSALGKSPMLAGNELTVADVLMGV 284  
Db 185 HNAVNATLIDSVWDIAIFQLKEGSSKEKAAVFRSNMSALGKSPMLAGNELTVADVLMGV 244  
Qy 285 LQOIGGCSVTVPANVQRMRSKCNLAPFTALKCLK 320  
Db 245 LQOIGGCSVTVPANVQRMRSKCNLAPFTALKCLK 280

## RESULT 7

Q6DKR6 PRELIMINARY; PRT; 311 AA.  
AC Q6DKR6; 25-OCT-2004 (TReMBLrel. 28, Created)  
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)  
DE 25-OCT-2004 (TReMBLrel. 28, Last annotation update)  
MGCG9221 protein.  
GN Name=MGCG9221;  
OS Xenopus tropicalis (western clawed frog) (Silurana tropicalis).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
OC Xenopodinae; Xenopus.  
NCBI\_TaxID=8364;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshlyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millany S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
Krzyszinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,  
Jones S.J., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RL Klein S., Gerhard D.S.;  
Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC074561; AAH74561.1; -;  
DR InterPro; IPR004046; GST\_Cterm.  
DR InterPro; IPR010987; GST\_C\_1like.  
DR Pfam; PF00043; GST\_C\_1.  
SQ SEQUENCE 311 AA; 34480 MW; D98F27F73C466154 CRC64;

Query Match 68.7%; Score 1144.5; DB 2; Length 311;  
Best Local Similarity 68.8%; Pred. No. 1.5e-84;  
Matches 220; Conservative 43; Mismatches 48; Indels 9; Gaps 3;

Qy 1 MEMYQVPRYHGGAPLRELPTCMYRLPNVHGRSYPAGAGHVOESNLSQALBSROD 60  
Db 1 MEMYQVPRYHGGAPLRELPTCMYRLPNVHGRSYPAGAGHVOESNLSQALBSROD 52  
Qy 61 DILKRLYELKAAVNDGSKITQTPDADVTNITLQADEPTLATNADLNSVLCGVKAL 120  
Db 53 DILKRLYELKAAVNDGSKITQTPDADVTNITLQADEPTLATNADLNSVLCGVKAL 112  
Qy 121 DIVINANPASPPLSLVHLHLCBPRVSLVTHHSSVKSPENLKKCFGEQKKQKQPROD 180  
Db 113 DIVINANPASPPLSLVHLHLCBPRVSLVTHHSSVKSPENLKKCFGEQKKQKQPROD 172  
Qy 181 YOLGFTLWKQVPTQMKFSIQTMCPLEGSGNIAFLFSLFGOKHNAVNATLIDSVWDIA 240  
Db 173 YOLGFTLWKQVPTQMKFSIQTMCPLEGSGNIAFLFSLFGOKHNAVNATLIDSVWDIA 232  
Qy 241 IFQLKEGSSKEKAAVFRSNMSALGKSPMLAGNELTVADVLMGVLQOIGGCSVTVPANQ 300  
Db 233 IFQLKEGSSKEKAAVFRSNMSALGKSPMLAGNELTVADVLMGVLQOIGGCSVTVPANQ 291  
Qy 301 RMRSCENLAPFTALKCLK 320  
Db 292 RMRSCENLAPFTALKCLK 311

## RESULT 8

Q6IND4 PRELIMINARY; PRT; 311 AA.  
AC Q6IND4; 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)  
MGCG80304 protein.  
GN Name=MGCG80304;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
OC Xenopodinae; Xenopus.  
NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshlyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millany S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
Krzyszinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,  
Jones S.J., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
Richardson P.;  
"Genetic and genomic tools for Xenopus research: The NIH Xenopus  
Dev. Dyn. 225:384-391(2002).  
RN [3]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC072178; AAH72178.1;  
 DR InterPro: IPR004046; GST\_Cterm.  
 DR InterPro: IPR010987; GST\_C\_1like.  
 DR Pfam: PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34424 MW; D5E8325C18D88751 CRC64;

Query Match 68.4%; Score 1139.5; DB 2; Length 311;  
 Best Local Similarity 69.1%; Pred. No. 3.8e-84;  
 Matches 221; Conservative 40; Mismatches 50; Indels 9; Gaps 3;

QY 1 MPMYQVRYHGGAPRLVELPTCMYRLPNVHGRSGYPAPAGHVOESNLSIQALSSRD 60  
 DB 1 MPMYRMQPYCGG--EIQVDLPCTMYRLPNVHAT-----AANHQQADPAIQALSSROE 52  
 QY 61 DILKRIYELKAAVDGSKMIOPTDADLVNTIIQADEPTTLTTNALDINSVLGKDYGALK 120  
 DB 53 DILKRIYELKAAVDGSKMIOPTDADLVNTIIQADEQSATGSVYADLVLDYGLALR 112  
 QY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSVSVSPENILKCGEQNKQPPROD 180  
 DB 113 DIVINANPSPSLVLRLLCEHFRVLSVTHSSVSVSPENILKCGEQNKQPPROD 172  
 QY 161 YOLGFTLTKVNPKTQMKFSIQTMCPLEGGINAFPLFSLGQKHNANATLIDSWDIA 240  
 DB 173 YOLGFTLTKVNPKTQMKFSIQTMCPLEGGINAFPLFSLGQKHNANATLIDSWDIA 232  
 QY 241 IFOLKESSSKERKAAYFRSMNSALGKSPMLAGNELTVADVILMSVLOQTGGCSVTYPANVQ 300  
 DB 233 IFOLKESSSKERKAAYLKKANNTLIGKSPMLVGNELTVADVILMSVCAVQCGN-STVVPNVQ 291  
 QY 301 RWRKSCENLAPPNTALIKLK 320  
 DB 292 KWRKSCENLASFYSVLRFLK 311

## RESULT 9

ID 07ZYD7 PRELIMINARY; PRT; 311 AA.  
 AC 07ZYD7.  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Jv1-prov protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M., Skalski U., Smalins D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Maitra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT Initiative";  
 RL Dev. Dyn. 225:384-391 (2002).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC043832; AAH43832.1;  
 DR InterPro: IPR004046; GST\_Cterm.  
 DR InterPro: IPR010987; GST\_C\_1like.  
 DR Pfam: PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34647 MW; 2F08C9DA60DD63BE CRC64;

Query Match 67.9%; Score 1132.5; DB 2; Length 311;  
 Best Local Similarity 69.1%; Pred. No. 1.4e-83;  
 Matches 221; Conservative 39; Mismatches 51; Indels 9; Gaps 4;

QY 1 MPMTQVRYHGGAPRLVELPTCMYRLPNVHGRSGYPAPAGHVOESNLSIQALSSRD 60  
 DB 1 MPMTQVRYHGGAPRLVELPTCMYRLPNVHGRSGYPAPAGHVOESNLSIQALSSRD 52  
 QY 61 DILKRIYELKAAVDGSKMIOPTDADLVNTIIQADEPTTLTTNALDINSVLGKDYGALK 120  
 DB 53 DILKRIYELKAAVDGSKMIOPTDADLVNTIIQADEHAPSGSVYADLVLDYGLALR 112  
 QY 121 DIVINANPASPPLSLVLRLLCEHFRVLSVTHSSVSVSPENILKCGEQNKQPPROD 180  
 DB 113 DIVINANPSPSLVLRLLCEHFRVLSVTHSSVSVSPENILKCGEQNKQPPROD 172  
 QY 161 YOLGFTLTKVNPKTQMKFSIQTMCPLEGGINAFPLFSLGQKHNANATLIDSWDIA 240  
 DB 173 YOLGFTLTKVNPKTQMKFSIQTMCPLEGGINAFPLFSLGQKHNANATLIDSWDIA 232  
 QY 241 IFOLKESSSKERKAAYFRSMNSALGKSPMLAGNELTVADVILMSVLOQTGGCSVTYPANVQ 300  
 DB 233 IFOLKESSSKERKAAYLKKANNTLIGKSPMLVGNELTVADVILMSVCAVQCGN-STVVPNVQ 291  
 QY 301 RWRKSCENLAPPNTALIKLK 320  
 DB 292 KWRKSCENLASFYSVLRFLK 311

## RESULT 10

ID 07T3C0 PRELIMINARY; PRT; 321 AA.  
 AC 07T3C0.  
 DT 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Zgc:63976.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RC	STRAIN=PEST:
RA	Anopheles Genome Sequencing Consortium;
RL	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
CC	-1- CAUTION: The sequence shown here is derived from an
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
DR	EMBL: AAAB01008960; EAA10870.1; -.
FT	NON_TBR 1 1
SQ	SEQUENCE 340 AA; 37467 MW; FD0DB8525CDAEB976 CRC64;
	Query Match 16.0%; Score 267.5; DB 2; Length 340;
	Best Local Similarity 29.2%; Pred. 2.6e-13;
	Matches 100; Conservative 47; Mismatches 133; Indels 63; Gaps 13
QY	3 MYGVKPYHGGGAPRLVLEPTCMYRLPNVHGSRYP-----APG---AGHV----- 44
DB	12 MYRLKPYM--NQDVCELPCTMYLTKLVCAVSHDPKAFSADATGTLAAAGVDPMSIID 69
QY	45 ---QEEENSLQALRESQDDILKRYLTKKAVDGLSKMIGTDPDLDV--TWIIQADEPTT 100
DB	70 NLLQKANEDELRLMLAEKQORVLOQLAEK-----KEIMARTELKLMNAPPAVQPS 122
QY	101 -----LTTNALDLNSVLGKDGALADIVYANPASPSSLVLYHRLGCEHFRVLSYH 153
DB	123 PLKSKAQLKAEPINLT-----CLDDFVNAPSEVPYPSLLAKLIMKRLNLQVECF 174
QY	154 THSSVYSVPE-----NLKCFGEONKKOPRODQLGFTLLIKNV--PKYONKFSIQTWCP 206
DB	175 THSTVPLTSEBALAFQNAVYASGTAANLNR---IKVTLLIKNVGAVTEMTSPSYVP 230
QY	207 IEEGENTARFLFSLFGQKFN---AVNATLIDSWYDIAIPLQKSSSEKKAVFPSNSAL 263
DB	231 ICEGVILRLTSLSCGSEFEFVEQODNVDVSDIACYLLLNKONVQRQOILRTLGAKL 290
QY	264 GKSPMLAGNELTVADVILWISVLQIGGCSYTPVNVQRMWRSC 306
DB	291 GKAAFGGADLSLDCIAFTSAVKV---QRTVAKDVNPNNMKC 330
RESULT 12	
Q7KIMS PRELIMINARY; PRT; 301 AA.	
AC	Q7KIMS;
DT	05-JUL-2004 (TEMBLrel. 27, Created)
DT	05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT	05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE	CG12304-PB.
GN	ORFName=CG12304;
OS	Drosophila melanogaster (fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RP	MEDLINE=20196006; Pubmed=10731132; DOI=10.1126/science.287.5461.2185;
RA	Adame M.D., Celniker S.E., Holt R.A., Evans C.A., Goeanay J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoekins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Branton R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA	Abriel J.F., Aghayani A., An H.J., Andrews-Flanckoch C., Baldwin D.,
RA	Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.V., Bouck J., Brokstein P., Brothier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA	dosden K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA	Grodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,





```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE003510; AAF49612.1; -.
DR IntAct; Q9YUR3; -.
DR Flybase; FBgn0036515; CG12304.
DR InterPro; IPR010987; GST_C_1like.
DR InterPro; IPR004046; GST_Cterm.
DR Pfam; PF00043; GST_C; 1.
DR Protein biosynthesis.
KW DOMAIN
FT 280 327
SQ SEQUENCE 334 AA; 36933 MW; B68FD70AE621990F CRC64;

Query Match 12.9%; Score 214.5; DB 1; Length 334;
Best Local Similarity 25.6%; Pred. No. 5.2e-09;
Matches 85; Conservative 50; Mismatches 120; Indels 77; Gaps 10;

QY 18 VELPTCMRLPNV-----HGRSYGAPAGAGHVQESNLSLQA-----54
Db 13 IKLPICMYPPLKNVSLAADSLASGSTSASTSASTSCKLEANRIDRTGRNATCALDLDS 72
QY 55 -----LESRODDILKRLYEELKAVDGLSKMIGTPDADLDVTNIIQADEPT 99
Db 73 LGRQIQRLKDDTASVAKRQEKVLKQLELKAQIGQIR-----AGLGVCG--KTFQHT 123
QY 100 TLTNALDLNLSVLGKDYALKDVIYANPASPPLSLVLHRLCEHFRVLSTVHTSSVK 159
Db 124 TAFQNG-----GIKEVPLQDVVINGHPNFIYPALLALKNAMRNLYTIDVKTFHTSMA 176
QY 160 SV-----PENLKCFCGEONKQQRQDYQLGFTLIWKNVPKTQMKFSIOTMCPTEGEGN 212
Db 177 DIGPARREFEANKLV--PVPALPK---ISVTLIKNCEHTEMLSSPTMYVPYIGEYN 230
QY 213 IARFLFSLFGQKHNANATL---IDSWVDIAIFOLKEGSSKEKAAVFRSNNLSALGKSPWL 269
Db 231 IIRYLGIVGPAEYREYEGSPLCNEIDLVDICYQLRGNTHTKIQVAMVRLLDKRLQKQYF 290
QY 270 AGNELTAVADVILMSVLQOIGCSVTYPANQR 301
Db 291 GGSQMSVADVGVYSSL-----IRMPAYTERK 315
```

Search completed: February 23, 2005, 13:57:12  
Job time : 61.9811 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:25:39 ; Search time 70.371 Seconds  
(without alignments)  
1758.725 Million cell updates/sec

Title: US-10-622-817-5

Perfect score: 1655  
Sequence: 1 MPVYQVKNPHGSGAPLRVEL.....RMKSGCNLAPEFALQLLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1464	88.5	320	7	AD58613	Ad58613 Human Pro
2	1464	88.5	320	8	ABM60843	Abm60843 Tumour-as
3	1436	86.8	312	2	AAW25776	AAW25776 JTV1 Prot
4	1436	86.8	312	8	ADR86551	ADR86551 1-312 am1
5	1232	74.4	272	7	ADCI0204	ADCI0204 Human NOV
6	1069	64.6	229	8	ADR86553	ADR86553 84-312 am
7	735	44.4	161	8	ADR86552	ADR86552 1-161 am1
8	248	15.0	51	5	ABP01502	Abp01502 Human CRF
9	216.5	13.1	334	4	AB862468	Ab862468 Drosophila
10	148	8.9	263	4	ABG23964	ABG23964 Novel hum
11	126.5	7.6	716	3	AA641411	AA641411 Arabidops
12	126.5	7.6	719	3	AA641410	AA641410 Arabidops
13	126.5	7.6	748	3	AA641409	AA641409 Arabidops
14	106	6.4	1512	4	AAU04349	AAU04349 Mammalian
15	106	6.4	1512	4	AAW78732	AAW78732 Human pro
16	106	6.4	1550	4	AAW79716	AAW79716 Human pro
17	104	6.2	980	8	ADN23683	ADN23683 Bacterial
18	102	6.2	1049	6	ABM67090	ABM67090 Phototrab
19	101.5	6.1	724	4	AA685046	AA685046 Shrimp wh
20	99	6.0	691	6	ABU16445	ABU16445 Protein e
21	98	5.9	243	4	AAU58158	AAU58158 Propionib
22	98	5.9	243	6	ABM54677	ABM54677 Propionib
23	98	5.9	644	6	ABJ18952	ABJ18952 Pathogen
24	98	5.9	690	6	AAU34345	AAU34345 Staphyloc
25	98	5.9	690	6	ABM71749	ABM71749 Staphyloc

26	98	5.9	985	8	ADN88622	Adn88622 Rat epid
27	98	5.9	997	8	ADN88620	Adn88620 Rat epid
28	98	5.9	1004	8	ADN88621	Adn88621 Rat epid
29	98	5.9	1013	8	ADN88594	Adn88594 Rat epid
30	98	5.9	1016	8	ADN88619	Adn88619 Rat epid
31	98	5.9	1032	8	ADN88618	Adn88618 Rat epid
32	96.5	5.8	1398	8	ABM83058	Abm83058 Human dia
33	96.5	5.8	1398	8	ABM83057	Abm83057 Human dia
34	96.5	5.8	1440	8	ADN03674	Adn03674 Antipeori
35	96.5	5.8	1440	8	ADP54124	Adp54124 Human PRO
36	96.5	5.8	1440	8	ADP23041	Adp23041 PRO polyP
37	96	5.8	2273	2	AA898811	AA898811 Eryalphe
38	95	5.7	722	8	AD820980	AD820980 Bacterial
39	94	5.7	436	5	AB854068	AB854068 Lactococc
40	93.5	5.6	1064	5	AAE25162	AAE25162 Herbicida
41	93.5	5.6	1064	5	ABG17694	ABG17694 Novel hum
42	93.5	5.6	1090	4	ABG18331	ABG18331 Novel hum
43	93	5.6	204	8	AD821772	AD821772 Bacterial
44	93	5.6	559	2	AA810682	AA810682 Polymydr
45	93	5.6	559	2	AA810682	AA810682 Polymydr

#### ALIGNMENTS

RESULT 1	AD58613	standard; protein; 320 AA.
ID	AD58613	
XX	AD58613;	
AC	AD58613;	
XX		
DT	29-JAN-2004	(first entry)
XX		
DE	Human Protein Q13155, SEQ ID NO 4489.	
XX		
KW	Human; pain; neuronal tissue; gene therapy;	
KM	spinal segmental nerve injury; chronic constriction injury; CCI;	
KW	spared nerve injury; SNI; Chung.	
XX		
OS	Homo sapiens.	
XX		
PN	WO2003016475-A2.	
XX		
PD	27-FEB-2003.	
XX		
PF	14-AUG-2002; 2002WO-US025765.	
XX		
PR	14-AUG-2001; 2001US-0312147P.	
PR	01-NOV-2001; 2001US-0346382P.	
PR	26-NOV-2001; 2001US-0333347P.	
XX		
PA	(GEHO ) GEN HOSPITAL CORP.	
PA	(FARB ) BAYER AG.	
XX		
PI	Woolf C, D'ureo D, Befort K, Costigan M;	
DR	WPI, 2003-268312/26.	
DR	GENBANK; Q13155.	
XX		
PT	New composition comprising two or more isolated polypeptides, useful for	
PT	preparing a medicament for treating pain in an animal.	
XX		
PS	Claim 1, Page; 1017pp; English.	
XX		
CC	The invention discloses a composition comprising two or more isolated rat	
CC	or human polynucleotides or a polynucleotide which represents a fragment,	
CC	derivative or allelic variation of the nucleic acid sequence. Also	
CC	claimed are a vector comprising the novel polynucleotide, a host cell	
CC	comprising the vector, a method for identifying a nucleotide sequence	
CC	which is differentially regulated in an animal subjected to pain and a	
CC	kit to perform the method, an array, a method for identifying an agent	
CC	that increases or decreases the expression of the polynucleotide sequence	
CC	that is differentially expressed in neuronal tissue of a first animal	

CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNII) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 320 AA;

Query Match 88.5%; Score 1464; DB 7; Length 320;  
Best Local Similarity 87.2%; Pred. No. 7.5e-142;

Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPADAGHVQETSEPSIQALESROD 60  
DB 1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPADAGHVQETSEPSIQALESROD 60  
QY 61 DILKRLVELKRAVDGSKMTQTPDADVDVNIIOADEPPTLATNTLDLNSVIGKDYGALK 120  
DB 61 DILKRLVELKRAVDGSKMTQTPDADVDVNIIOADEPPTLATNTLDLNSVIGKDYGALK 120  
QY 121 DIVINANPASPPLSLVLRHLCEHRYVSTVHTHSSVGNVENVKCGEQAROSRHE 180  
DB 121 DIVINANPASPPLSLVLRHLCEHRYVSTVHTHSSVGNVENVKCGEQAROSRHE 180  
QY 181 YQGFETIMKNVPRKTKSVQVMCIIEBGNIAFLFSLFQOKNNAVTLTLDISWVDA 240  
DB 181 YQGFETIMKNVPRKTKSVQVMCIIEBGNIAFLFSLFQOKNNAVTLTLDISWVDA 240  
QY 241 MFOUREGSSKEKAVERSMNSALGRSPMLVGNELTVADVILMSVQCGGSSGAAPTVO 300  
DB 241 MFOUREGSSKEKAVERSMNSALGRSPMLVGNELTVADVILMSVQCGGSSGAAPTVO 300  
QY 301 RMLKSCENLAPFSTALQILK 320  
DB 301 RMLKSCENLAPFSTALQILK 320

## RESULT 2

ABM80843  
ID ABM80843 standard; protein; 320 AA.

XX  
AC ABM80843;

XX  
DT 18-NOV-2004 (first entry)

XX  
DE Tumour-associated antigenic target (TAT) polypeptide PRO81501, SEQ:2179.

XX  
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
XX  
XX gene therapy; cytosolic.

OS Homo sapiens.

XX  
PN WO2004030615-A2.

XX  
PD 15-APR-2004.

XX  
PF 29-SEP-2003; 2003WO-US028547.

XX  
PR 02-OCT-2002; 2002US-0414971P.

XX  
PA (GETH ) GENENTECH INC.

XX  
PI Wu TD, Zhang Z, Zhou Y;

XX  
DR WPI; 2004-347921/32.

XX  
N-PSDB; ACN38579.

PT New tumour-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.

XX  
PS Claim 12; SEQ ID NO 2179; 7273pp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
XX represents a TAT polypeptide of the invention

XX  
SQ Sequence 320 AA;

Query Match 88.5%; Score 1464; DB 8; Length 320;  
Best Local Similarity 87.2%; Pred. No. 7.5e-142;

Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPADAGHVQETSEPSIQALESROD 60  
DB 1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPADAGHVQETSEPSIQALESROD 60  
QY 61 DILKRLVELKRAVDGSKMTQTPDADVDVNIIOADEPPTLATNTLDLNSVIGKDYGALK 120  
DB 61 DILKRLVELKRAVDGSKMTQTPDADVDVNIIOADEPPTLATNTLDLNSVIGKDYGALK 120  
QY 121 DIVINANPASPPLSLVLRHLCEHRYVSTVHTHSSVGNVENVKCGEQAROSRHE 180  
DB 121 DIVINANPASPPLSLVLRHLCEHRYVSTVHTHSSVGNVENVKCGEQAROSRHE 180  
QY 181 YQGFETIMKNVPRKTKSVQVMCIIEBGNIAFLFSLFQOKNNAVTLTLDISWVDA 240  
DB 181 YQGFETIMKNVPRKTKSVQVMCIIEBGNIAFLFSLFQOKNNAVTLTLDISWVDA 240  
QY 241 MFOUREGSSKEKAVERSMNSALGRSPMLVGNELTVADVILMSVQCGGSSGAAPTVO 300  
DB 241 MFOUREGSSKEKAVERSMNSALGRSPMLVGNELTVADVILMSVQCGGSSGAAPTVO 300  
QY 301 RMLKSCENLAPFSTALQILK 320  
DB 301 RMLKSCENLAPFSTALQILK 320

## RESULT 3

AAM25776

ID	AAW25776	standard; protein; 312 AA.
AC	AAW25776;	
XX		
DT	19-DEC-1997	(first entry)
XX		
DE	JTV1 protein.	
XX		
KM	JTV1; hPMS2; probe; detection; chromosome 7; deletion;	
KW	mismatch repair gene; hereditary non-polyposis colorectal cancer;	
XX	homologous recombination.	
OS	Homo sapiens.	
XX		
FN	W09708312-A1.	
XX		
PD	06-MAR-1997.	
XX		
PE	26-AUG-1996; 96MO-US013598.	
XX		
PR	24-AUG-1995; 95US-00518862.	
XX		
PA	(UXXJO ) UNIV JOHNS HOPKINS.	
XX		
PI	Vogelstein B, Kinzler KW, Nicolaides NC;	
XX		
DR	WPI; 1997-179269/16.	
DR	N-PSDB; AAT86182.	
XX		
PT	Novel chromosome 7 gene, JTV1 - used for detecting chromosome 7	
XX	deletions, and PMS2 promoter activity.	
XX		
PS	Claim 5; Fig 2; 31pp; English.	
XX		
CC	This sequence is JTV1 protein and is encoded by DNA isolated from human	
CC	chromosome 7. The JTV1 coding sequence is located upstream from hPMS2.	
CC	JTV1 cDNA can be used as probes to detect chromosome 7 deletions	
CC	involving JTV1. Due to the overlapping promoter regions, deletions of	
CC	JTV1 would also affect PMS2 (a mismatch repair gene) expression, leading	
CC	to hereditary non-polyposis colorectal cancer. JTV1 can also be used to	
CC	assay activity or competence of the PMS2 promoter region. The presence of	
CC	JTV1 suggesting that the PMS2 promoter is intact. JTV1 sequences can also	
CC	be used to guide homologous recombination at the PMS2 locus	
XX		
CC		
XX		
SQ	Sequence 312 AA;	
Query Match	86.8%; Score 1436; DB 2; Length 312;	
Best Local Similarity	87.5%; Pred. No. 5,6e-139;	
Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;		
QY	1 MPMYVVKRYHGGGSAFLRYELPTCMKRLPNVHSHKTTSPATDAGHVQETSEPSIQALESROD 60	
DB	1 MPMYVVKRYHGGGSAFLRYELPTCMKRLPNVHSHKTTSPATDAGHVQETSEPSIQALESROD 60	
QY	61 DILKRLYEILKAADVGLSKMIMHTPDADLDVNTILQADEPTTLATNTLDINSVIGKDYGALK 120	
DB	61 DILKRLYEILKAADVGLSKMIMHTPDADLDVNTILQADEPTTLATNTLDINSVIGKDYGALK 120	
QY	121 DIVINANPASPLSLVLVHRLICEKRYVLSTYHTHSSVKVNPENLVKCFGEQARKQSHK 180	
DB	121 DIVINANPASPLSLVLVHRLICEKRYVLSTYHTHSSVKVNPENLVKCFGEQARKQSHK 180	
QY	181 YOLGFTLLIMKQVPTKQMPKSVOTMCPRIEGBGNIAFLPSLFCQKNAATVTLTIDISWVIA 240	
DB	181 YOLGFTLLIMKQVPTKQMPKSVOTMCPRIEGBGNIAFLPSLFCQKNAATVTLTIDISWVIA 240	
QY	241 MFOLEGGSSKEKAAVFRSNMNLGKSPMLVGNELTVADVVLMSVLAQQTGGSSGAATVNVQ 300	
DB	241 IFOLKEGSSKEKAAVFRSNMNLGKSPMLVGNELTVADVVLMSVLAQQTGGSSGAATVNVQ 300	
QY	301 RWLKSCENLAPF 312	
DB	301 RWLKSCENLAPF 312	

RESULT 4	
ID ADR86551	
AC ADR86551 standard; protein, 312 AA.	
XX	
XX ADR86551;	
XX	
DT 18-NOV-2004 (first entry)	
XX	
DE 1-312 amino acid sequence of p38/JTV-1 protein.	
XX	
XX p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.	
XX	
KW Homo sapiens.	
OS	
PN EP1454628-A2.	
PD	
PD 08-SEP-2004.	
PF	
PF 09-SEP-2003; 2003EP-00020344.	
PR	
PR 03-MAR-2003; 2003KR-00013058.	
PA (UNSE-) UNIV SEOUL NAT IND FOUND.	
PI Kim S, Park B;	
PI N-PIDB; ADR86548.	
DR WPI; 2004-627822/61.	
XX	
PT New isolated p38/JTV-1 protein, useful as medicament for treating cancer	
PT e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer	
PT and cutaneous or intraocular melanoma, as well as for screening new	
PT anticancer agents.	
PS Claim 5; SEQ ID NO 4; 47pp; English.	
XX	
XX The present invention relates to an isolated p38/JTV-1 protein for use as	
CC medicament. The p38/JTV-1 protein or the pharmaceutical composition is	
CC useful as medicament for treating breast cancer, large intestinal cancer,	
CC lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood	
CC cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,	
CC cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,	
CC rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,	
CC endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,	
CC Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine	
CC cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue	
CC tumour, uethral cancer, penile cancer, prostate cancer, chronic or acute	
CC leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter	
CC cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary	
CC CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary	
CC adenoma, or their combination. The protein is useful as a target for	
CC screening new anticancer agents. The present sequence represents the 1-	
CC 312 amino acid sequence of p38/JTV-1 protein.	
XX	
SEQ Sequence 312 AA;	
QY Query Match 86.8%; Score 1436; DB 8; Length 312;	
DB Best Local Similarity 87.5%; Pred. No. 5.66-139;	
Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;	
QY 1 MPMVQVRKHGGSAPLRYELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSLOALESROD 60	
DB 1 MPMVQVRKHGGAPLRVELPTCMYRLPNVHGSRYPGAPGHVQESNNLSLOALSHROD 60	
QY 61 DILKLRYELKAAVDGLSKMITHPTPADLDVTNIIQADEFITLATNTLDINSVLGXDYGALK 120	
DB 61 DILKLRYELKAADVGLSKMIQTTPADDDVTNIIQADEFITLTNALDLNSVLGDYGLXK 120	
QY 121 DIVINANPASPPLSLVTHRLLCERFRLSTVTHSSSVKNVPENLVKCFGQAKKQSHE 180	
DB 121 DIVINANPASPPLSLVTHRLLCERFRLSTVTHSSSVKSPENLVKCFGQONKKOPROD 180	

```

QY 181 YQGLTILIMKNVPTOMKFSVOTMCPTEGEGNIAFLFSLFGOKNAVTLTIDSWDIA 240
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 181 YQGLTILIMKNVPTOMKFSVOTMCPTEGEGNIAFLFSLFGOKNAVTLTIDSWDIA 240
Db 241 MFQLRGSSSKKKAAPFRSMNSALGRSPMLVGNELTVADVVLMSVLCQGTGSSGAAPTIVQ 300
QY 241 IFQLEKSSSKKKAAPFRSMNSALGRSPMLAGNELTVADVVLMSVLCQGTGSSVTPANVQ 300
Db 301 RMLKSCENLAPF 312
QY 301 RMLKSCENLAPF 312
Db 301 RMLKSCENLAPF 312

```

```

RESULT 5
ADCI0204
ID ADCl0204 standard; protein; 272 AA.
XX
AC ADCl0204;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human NOVX polypeptide SEQ ID NO: 226.
XX
KW cytosolic; antidiabetic; anorectic; cerebroprotective; neuroprotective;
KW antiinflammatory; gene therapy; antisense therapy; thymomimetic; NOVX;
KW pathology; cancer; diabetes; obesity; endocrine disorder; CNS disorder;
KW inflammatory disorder; chromosome mapping; tissue typing;
KW predictive medicine.
XX
OS Homo sapiens.
XX
PN WO200300842-A2.
XX
PD 03-JAN-2003.
XX
PF 04-JUN-2002; 2002WO-US017443.
XX
PR 04-JUN-2001; 2001US-0295607P.
PR 04-JUN-2001; 2001US-029561P.
PR 06-JUN-2001; 2001US-029640P.
PR 06-JUN-2001; 2001US-029641P.
PR 07-JUN-2001; 2001US-029657P.
PR 11-JUN-2001; 2001US-029741P.
PR 12-JUN-2001; 2001US-0295573P.
PR 12-JUN-2001; 2001US-0297567P.
PR 14-JUN-2001; 2001US-0298285P.
PR 15-JUN-2001; 2001US-0298528P.
PR 18-JUN-2001; 2001US-0299133P.
PR 19-JUN-2001; 2001US-0299230P.
PR 21-JUN-2001; 2001US-029949P.
PR 22-JUN-2001; 2001US-0300177P.
PR 26-JUN-2001; 2001US-0300883P.
PR 28-JUN-2001; 2001US-0301530P.
PR 28-JUN-2001; 2001US-030150P.
PR 03-JUL-2001; 2001US-0302851P.
PR 31-JUL-2001; 2001US-030890P.
PR 14-SEP-2001; 2001US-032297P.
PR 25-SEP-2001; 2001US-032469P.
PR 03-DEC-2001; 2001US-0337477P.
PR 14-DEC-2001; 2001US-0341562P.
PR 21-FEB-2002; 2002US-038656P.
PR 21-FEB-2002; 2002US-038656P.
PR 22-FEB-2002; 2002US-0359034P.
PR 22-FEB-2002; 2002US-0359035P.
PR 22-FEB-2002; 2002US-0359122P.
PR 22-FEB-2002; 2002US-0358978P.
PR 22-FEB-2002; 2002US-0359034P.
PR 27-FEB-2002; 2002US-0359121P.
PR 01-MAR-2002; 2002US-0358964P.
PR 12-MAR-2002; 2002US-0360858P.
PR 12-MAR-2002; 2002US-0363430P.
PR 10-APR-2002; 2002US-0363676P.
PR 10-APR-2002; 2002US-0371346P.
PR 10-MAY-2002; 2002US-0379444P.
XX

```

```

PR 04-JUN-2002; 2002US-00379444.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Agee ML, Anderson DW, Berghs C, Caeman SJ, Catterton E;
PI Delpo VA, Edinger SR, Eisen A, Ellemann K, Gangoli EA;
PI Gerlach VL, Gorman U, Guo X, Herrmann JM, Hjalte T, Ji W, Kekuda R;
PI Khrantsov NV, Li L, Liu X, Malyanar UM, Miller CE, Miller I;
PI Ott T, Padigaru M, Patturajan M, Pena CE, Rastelli L, Rieger DK;
PI Rothenberg ME, Shenoy SG, Shimkels RA, Smithson G, Spaderna SK;
PI Spletke KA, Stone DJ, Vernet CAM, Zhong H, Zhong W, Alsbrook JP;
PI Burgess CE, Lepley DM;
XX
DR WPI; 2003-210149/20.
DR N-PSDB; ADCl0203.
XX
PT New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
XX Claim 1; SEQ ID NO 226; 772bp; English.
XX
PS The invention relates to novel isolated polypeptides, mature form of the
CC polypeptide, a sequence that is 95% identical to the polypeptide or the
CC polypeptide comprising one or more conservative substitutions. The NOVX
CC polypeptide is useful for treating or preventing a pathology associated
CC with the polypeptide e.g. disorders associated with aberrant expression
CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC endocrine, CNS and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. This sequence corresponds to one of the
CC polypeptides of the invention.
XX
SQ Sequence 272 AA;

```

```

Query Match 74.4%; Score 1232; DB 7; Length 272;
Best Local Similarity 88.4%; Pred. No. 5,1e-118;
Matches 237; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 45 QETSEPSIQLESQDDILRLVELKAAVDGLSKMTHTPDADVDVNILOADEPTLATN 104
Db 5 QESNLSIQLESQDDILRLVELKAAVDGLSKMTHTPDADVDVNILOADEPTLATN 64
QY 105 TLIDNSVIGKDYGLKDVINANPASPPLSLVHLHCERYVSLVSTVTHSSVKNVPE 164
Db 65 ALDINSVIGKDYGLKDVINANPASPPLSLVHLHCERYVSLVSTVTHSSVKNVPE 124
QY 165 LVKCFGQARKQSRHVEYQGLFTLIMKNVPTOMKFSVOTMCPTEGEGNIAFLFSLFGOK 224
Db 125 LKCFGEQNKQPRQDYQGLFTLIMKNVPTOMKFSVOTMCPTEGEGNIAFLFSLFGOK 184
QY 225 HNAVTLTIDSWDIAIMFOLARESSSKKKAAPFRSMNSALGRSPMLVGNELTVADVVLMSV 284
Db 185 HNAVTLTIDSWDIAIMFOLARESSSKKKAAPFRSMNSALGRSPMLVGNELTVADVVLMSV 244
QY 285 LQGTGSSGAAPTIVQRLKSCENLAPF 312
Db 245 LQGTGSSGAAPTIVQRLKSCENLAPF 272

```

```

RESULT 6
ADRI6553
ID ADRI6553 standard; protein; 229 AA.
XX
AC ADRI6553;
XX
DT 18-NOV-2004 (first entry)
XX
DE 84-312 amino acid sequence of p38/JTV-1 protein.
XX
KW p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.
XX

```

DE		1-161 amino acid sequence of p38/JTV-1 protein.
XX		
KM	p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.	
XX		
OS	Homo sapiens.	
XX		
PN	EPI454628-A2.	
XX		
PD	08-SEP-2004.	
XX		
PF	09-SEP-2003; 2003EP-00020344.	
XX		
PR	03-MAR-2003; 2003KR-00013058.	
FA	(UYSE-) UNIV SEOUL NAT IND FOUND.	
XX		
PI	Kim S, Park B;	
DR	WPI: 2004-627822/61.	
XX	N-PSDB; ADR86549.	
PT	New isolated p38/JTV-1 protein, useful as medicament for treating cancer	
XX	e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer	
PT	and cutaneous or intraocular melanoma, as well as for screening new	
XX	anticancer agents.	
ET		
XX		
PS	Claim 5; SEQ ID NO 5; 47bp; English.	
XX		
CC	The present invention relates to an isolated p38/JTV-1 protein for use as	
CC	medicament. The p38/JTV-1 protein or the pharmaceutical composition is	
CC	useful as medicament for treating breast cancer, large intestinal cancer,	
CC	lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood	
CC	cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,	
CC	cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,	
CC	rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,	
CC	endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,	
CC	Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine	
CC	cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue	
CC	tumour, urethral cancer, penile cancer, prostate cancer, chronic or acute	
CC	leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter	
CC	cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary	
CC	CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary	
CC	adenoma, or their combination. The protein is useful as a target for	
CC	screening new anticancer agents. The present sequence represents the 1-	
CC	161 amino acid sequence of p38/JTV-1 protein.	
XX		
SQ	Sequence 161 AA:	
	Query Match 44.4%; Score 735; DB 8; Length 161;	
	Best local Similarity 80.8%; Pred. No. 4,4e-67;	
	Matches 143; Conservative % 5; Indels 13; Gaps 0;	
OY	1 MPWMGVRRYGSGAPLRLVELPTCMYRPNVHAKTKTSPATDAGHQETSBPGLAESROD 60	
DB	1 MPMGVRRYHGGAAPLRLVELPTCMYRPNVHGRGYPGAGAHQEBSNNLSLOLESROD 60	
OY	61 DILKRLLELAANDGLSKMITPPADDDVTNIIOADEPPTLTATTLDLNSVLGKDYGALK 120	
DB	61 DILKRLYLEKAANDGLSKMITPPADDVTNIIOADEPPTLTNTAALDNSVLGKDYGALK 120	
OY	121 DIVIANPASPPSLVLVLRLLCERYAVLSTVTHSSVKVNV 161	
DB	121 DIVIANPASPPSLVLVLRLLCEHPFVLSTVTHSSVKSV 161	
RESULT 8		
ABP01502		
ID	ABP01502 standard; protein; 51 AA.	
XX	ABP01502;	
AC		
XX		
DT	24-JUN-2002 (first entry)	
XX		

DE Human ORFX protein sequence SEQ ID NO:2986.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN MO200192523-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 29-MAY-2001; 2001WO-US010836.  
XX  
PR 30-MAY-2000; 2000US-0206132P.  
XX  
PR 29-AUG-2000; 2000US-0228716P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach MD;  
XX  
DR WPI; 2002-106308/14.  
XX  
DR N-PSDB; ABN17254.  
XX  
PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
PS Disclosure; SEQ ID NO 2986; 1037bp; English.  
XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX.  
CC Proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 51 AA;  
Query Match 15.0%; Score 248; DB 5; Length 51;  
Best Local Similarity 96.1%; Pred. No. 1.3e-17;  
Matches 49; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

74 DGLSKMIRHPDADLVNTNIIQADEPTTLATNTLINSVIGKDYGLAKDIOVI 124  
|||||  
1 DGLSKMIRHPDADLVNTNIIQADEPTTLATNTLINSVIGKDYGLAKDIOVI 51

RESULT 9  
ABB62468

ID ABB62468 standard; protein; 334 AA.  
XX  
AC ABB62468;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 14196.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN MO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL06571.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 14196; 21bp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins (AB57737-  
CC ABH22072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 334 AA;  
Query Match 13.1%; Score 216.5; DB 4; Length 334;  
Best Local Similarity 26.0%; Pred. No. 4.8e-13;  
Matches 86; Conservative 48; Mismatches 124; Indels 73; Gaps 9;

QY 3 MYQVKPRHGSAPRLVRLPTCMYRLPNV-----HSKTTSPATDAG----- 42  
|||  
1 MYELKTL--LPQFDIKLPTCMYRLKNVSLAASGTSSTASTSSTSCLEANNRID 57  
|||  
QY 43 -----HVQETSSEPSLOALRESRODILKRLVELKAADVGSKMIHTPD 84  
|||  
58 RTGNMATCALDDSDLSGRQGLLKDDPYASVARQEKVTLKQLELKAQGLQIR----- 110  
|||  
QY 85 ADDLVNTIIQADEPTTLATNTLINSVIGKDYGLAKDIOVINANPASPISLVTHRLCB 144  
|||  
111 AGLGVCG--KTFOHTTAFQNG-----GLKEVPLQDVINGHPNFIYPALLALKNAMRN 161  
|||  
QY 145 RYRVLSVTHHSSVKNV-----PENIVKCSBEQARKSRHEYQGLFTLIMKNVFKTQM 197  
|||  
162 LVTIIDVKTFTFHSYMAIDIGPAAREFEANLAKVPVNPPLP-----KISVTLIMKNCENTEM 215  
|||  
QY 198 KFSVQTMCPRIEGBGNIAEFSLFGQKHNAVLTTL---IDSWVDLAMPOLREGSSKEKA 254  
|||  
216 ISSPTMYVPIYGEVNIIRIVLGRGPAEYRVEBSPLCNEIDVLDICYQLLRCHTKHTQYA 275  
|||  
QY 255 VFRSNNASALGRSPWLVGNELTVADVLMGVL 285

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:36:20 ; Search time 17.81 Seconds  
(without alignment)  
1341.256 Million cell updates/sec

Title: US-10-622-817-5  
Perfect score: 1655  
Sequence: 1 MPWQVKPHGSAPLRVEL.....RWLKSCEMLAPFSTALQLLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCOTS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1464	88.5	341	4 US-09-949-016-11312	Sequence 11312, A
2	1436	86.8	312	2 US-08-518-862C-2	Sequence 2, Appl1
3	106	6.4	1512	3 US-09-443-184-48	Sequence 48, Appl
4	96.5	5.8	1440	3 US-09-357-251-37	Sequence 37, Appl
5	90.5	5.5	359	3 US-09-540-824-2	Sequence 2, Appl1
6	90	5.4	559	2 US-08-756-317-10	Sequence 10, Appl
7	90	5.4	559	4 US-09-091-609-4	Sequence 4, Appl1
8	90	5.4	2954	4 US-09-150-867-1	Sequence 1, Appl1
9	90	5.4	5215	4 US-09-105-537-2	Sequence 2, Appl1
10	89	5.4	1056	4 US-09-595-684B-29	Sequence 29, Appl
11	89	5.4	1057	3 US-09-541-782-10	Sequence 10, Appl
12	89	5.4	1057	4 US-09-723-820-10	Sequence 10, Appl
13	89	5.4	1057	4 US-10-270-085-10	Sequence 10, Appl
14	89	5.4	1287	4 US-09-949-016-7826	Sequence 7826, Ap
15	88	5.3	712	4 US-09-489-039A-10736	Sequence 10736, A
16	88	5.3	778	4 US-09-583-110-3930	Sequence 3930, Ap
17	88	5.3	786	4 US-09-107-433-3893	Sequence 3893, Ap
18	86.5	5.2	520	4 US-09-248-796A-17644	Sequence 17644, A
19	86.5	5.2	967	4 US-09-540-236-2449	Sequence 2449, Ap
20	86	5.2	535	4 US-09-489-039A-11461	Sequence 11461, A
21	85.5	5.2	556	4 US-09-134-000C-6329	Sequence 6329, Ap
22	85.5	5.2	608	4 US-09-284-768A-4	Sequence 4, Appl1
23	85.5	5.2	657	4 US-09-284-768A-7	Sequence 7, Appl1
24	85	5.1	222	4 US-09-248-796A-18146	Sequence 18146, A
25	85	5.1	302	4 US-09-107-532A-6924	Sequence 6924, Ap
26	85	5.1	559	4 US-09-672-749-2	Sequence 2, Appl1
27	85	5.1	1971	4 US-09-914-272A-1	Sequence 1, Appl1

28	85	5.1	1971	4 US-10-638-333-1	Sequence 1, Appl1
29	84	5.1	443	4 US-09-949-016-10598	Sequence 10598, A
30	84	5.1	646	4 US-09-248-796A-18678	Sequence 18678, A
31	84	5.1	1057	4 US-09-428-156B-2	Sequence 2, Appl1
32	83.5	5.0	360	4 US-09-949-016-10589	Sequence 10589, A
33	83.5	5.0	475	4 US-09-328-352-5379	Sequence 5379, Ap
34	83.5	5.0	515	4 US-09-489-039A-8208	Sequence 8208, Ap
35	83.5	5.0	519	4 US-09-543-681A-6505	Sequence 6505, Ap
36	83.5	5.0	789	3 US-09-002-285-84	Sequence 84, Appl
37	83.5	5.0	789	4 US-09-589-477-84	Sequence 84, Appl
38	83.5	5.0	789	4 US-10-099-285A-84	Sequence 84, Appl
39	83.5	5.0	959	4 US-09-543-681A-6879	Sequence 6879, Ap
40	83	5.0	434	3 US-09-012-072-4	Sequence 4, Appl1
41	83	5.0	434	3 US-09-120-601-4	Sequence 4, Appl1
42	83	5.0	448	3 US-09-120-601-6	Sequence 6, Appl1
43	83	5.0	559	4 US-09-821-016-1	Sequence 1, Appl1
44	83	5.0	559	4 US-10-266-787-1	Sequence 1, Appl1
45	83	5.0	767	3 US-08-836-567-8	Sequence 8, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-949-016-11312
; Sequence 11312, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 11312
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11312

Query Match      88.5%; Score 1464; DB 4; Length 341;
Best Local Similarity 87.2%; Pred. No. 9e+153;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY      1 MPWQVKPHGSAPLRVELPTCMYRLPNVHSKTTSPATDAGHVOFTSPBSLQALSSROD 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      22 MPWQVKPHGSAPLRVELPTCMYRLPNVHSKTTSPATDAGHVOFTSPBSLQALSSROD 81

QY      61 DLKRLYEKAAVNDGSKMHTPPADIDVTNTLQADEPTLTNTLLNSVSKDVGALK 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      82 DLKRLYEKAAVNDGSKMHTPPADIDVTNTLQADEPTLTNTLLNSVSKDVGALK 141

QY      121 DIVIANPASPPLSLVLRHLRCERYRVSTYHTSSVKNVDENLVKCFGEQAKRSRHE 180
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      142 DIVIANPASPPLSLVLRHLRCERYRVSTYHTSSVKNVDENLVKCFGEQAKRSRHE 201

QY      181 YOLGFTLIWKVNPXTQMKFSVOTMCPBIEGEGNIARFLFSLFGQKNAVTLTLIDSVNDA 240
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      202 YOLGFTLIWKVNPXTQMKFSVOTMCPBIEGEGNIARFLFSLFGQKNAVTLTLIDSVNDA 261

QY      241 MPQLREGSKKEKAAVFRSNMSALGSPMLVGNELTVADVNVSVYLOOTGSSGAATPNVQ 300
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      262 IPQLREGSKKEKAAVFRSNMSALGSPMLVGNELTVADVNVSVYLOOTGSSGAATPNVQ 321

QY      301 RWLKSCEMLAPFSTALQLLK 320
```

Db 322 RMWRSCENTAPFNTALKLK 341

RESULT 2  
US-08-518-862C-2  
; Sequence 2, Application US/08518862C  
; Patent No. 7,642,000

Query Match	86.8%;	Score 1436;	DB 2;	Length 312;
Best Local Similarity	87.5%;	Pred. No. 9.6e-150;		
Matches 273;	Conservative 14;	Mismatches 25;	Indels 0;	Gaps 0

QY	MEMVOVPHYGGASPIPLVELPTCMYLLPNVHSTKTTSPATDAGHVOETSEPSLOALIESROD	60
Db	1 MEMVOVPHYGGASPIPLVELPTCMYLLPNVHSTKTTSPATDAGHVOETSEPSLOALIESROD	60
QY	61 DILRLVLELKAANDGSLKMIHTPDDADLTNTNIIQADEPTTLATNTLDNSTVIGKDYALK	120
Db	61 DILRLVLELKAANDGSLKMIHTPDDADLTNTNIIQADEPTTLATNTLDNSTVIGKDYALK	120
QY	121 DIVANMAPSPSLIYVHRLLCERRYLSTYHTHTSSVXKYNPENLVKCFGEQARKOSRHE	180
Db	121 DIVANMAPSPSLIYVHRLLCERRYLSTYHTHTSSVXKYNPENLVKCFGEQARKOSRHE	180
QY	121 DIVANMAPSPSLIYVHRLLCERRYLSTYHTHTSSVXKYNPENLVKCFGEQARKOSRHE	180
Db	121 DIVANMAPSPSLIYVHRLLCERRYLSTYHTHTSSVXKYNPENLVKCFGEQARKOSRHE	180
QY	181 YOLGFTLLIMNVKOTMKFSSVOTMCPIBEGBGNIAFLSLTGCKHNAVNTLLIDSWDIA	240
Db	181 YOLGFTLLIMNVKOTMKFSSVOTMCPIBEGBGNIAFLSLTGCKHNAVNTLLIDSWDIA	240
QY	181 YOLGFTLLIMNVKOTMKFSSVOTMCPIBEGBGNIAFLSLTGCKHNAVNTLLIDSWDIA	240
Db	181 YOLGFTLLIMNVKOTMKFSSVOTMCPIBEGBGNIAFLSLTGCKHNAVNTLLIDSWDIA	240
QY	241 MPOLRSSSEKKAAYRSMNSALGSSPMLVGNELTVADVTIMSYLQOTGSSGAAPNTVQ	300
Db	241 MPOLRSSSEKKAAYRSMNSALGSSPMLVGNELTVADVTIMSYLQOTGSSGAAPNTVQ	300
QY	241 MPOLRSSSEKKAAYRSMNSALGSSPMLVGNELTVADVTIMSYLQOTGSSGAAPNTVQ	300
Db	241 MPOLRSSSEKKAAYRSMNSALGSSPMLVGNELTVADVTIMSYLQOTGSSGAAPNTVQ	300
QY	301 RMLKSCENTLAPF 312	
Db	301 RMLKSCENTLAPF 312	
QY	301 RMLKSCENTLAPF 312	
Db	301 RMLKSCENTLAPF 312	

RESULT 3  
US-09-443-184-48  
; Sequence 48, Application US/09443184A

```

; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1
US-09-443-184-48

```

Query Match	6.4%	Score 106	DB 3	Length 1512
Best Local Similarity	21.8%	Pred. No. 0.096		
Matches	44	Conservative 33	Mismatches 69	Indels 56
			Gaps	8
QY	124	INANASPEPSLILVHLRLCERYRLSVTHSSVKNVPENLVYCFGEQARKSRHHYQL	163	
Db	6	LTVNSGDPPLGALL-----AVEHYKQD-VSISVEGKENLH----	41	
QY	184	GFTLIWKVNPPTQMKFSVQTMCPLEGEGNIAFL-----FSLEFGQKHAVTLLTIDSWY	237	
Db	42	-----VSENVITTDV-----NSILIRYARVATTAAGLYGS--NIMEHTEIDHWL	82	
QY	238	DIAMFOLEGGSSKEKAAVFRSMNSALGRSPMLVGENELTVADVYLMSVLO-----QTGGSS	292	
Db	83	EFSAATKL--SSCDBFTSTINEINHCLSTRITLVGNSLSLADLCVWATLTKGNMAAMQEOLOKQ	140	
QY	293	GAAPNTVQRMKLSCENLAPST	314	
Db	141	KKAPVHVWRMGFLAQAQAFQS	162	

US-09-357-251-37

SEQUENCE 37 Application US/09357251  
Patent No. 6271441

GENERAL INFORMATION:

APPLICANT: Falco, S. Carl  
APPLICANT: Famodu, Layo O.  
APPLICANT: Orozco, Buddy  
APPLICANT: Schwabner, James S.  
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetases  
FILE REFERENCE: BB-11193  
CURRENT APPLICATION NUMBER: US/09/357,251  
CURRENT FILING DATE: 1999-07-20  
EARLIER APPLICATION NUMBER: 60/093,530  
EARLIER FILING DATE: July 21, 1998  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 37  
LENGTH: 1440  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-357-251-37



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:57:26 ; Search time 50.2443 Seconds  
(without alignments)  
2084.158 Million cell updates/sec

Title: US-10-622-817-5  
Perfect score: 1655  
Sequence: 1 MEMYGVKPYHGASAPLRYEL.....RWLKSCENLAPFSTALQLLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA.\*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1436	86.8	312	US-10-463-676-4	Sequence 4, Appl1
2	1069	64.6	229	US-10-463-676-6	Sequence 6, Appl1
3	735	44.4	161	US-10-463-676-5	Sequence 5, Appl1
4	105.5	6.4	925	US-10-437-963-126132	Sequence 126132,
5	104	6.3	980	US-10-369-493-6336	Sequence 6336, Ap
6	103.5	6.3	468	US-10-424-599-260388	Sequence 260388,
7	99.5	6.0	914	US-10-437-963-153870	Sequence 153870,
8	99	6.0	691	US-10-282-1122A-43659	Sequence 43659, A
9	98	5.9	690	US-09-815-242-5841	Sequence 5841, Ap
10	96.5	5.8	1440	US-09-864-464-37	Sequence 37, Appl
11	96.5	5.8	1440	US-10-786-720-31	Sequence 31, Appl
12	96	5.8	569	US-10-437-963-123904	Sequence 123904,
13	95	5.7	722	US-10-369-493-10013	Sequence 10013, A

14	93.5	5.6	456	15	US-10-425-114-63166	Sequence 63166, A
15	93	5.6	204	15	US-10-369-493-10805	Sequence 10805, A
16	93	5.6	1398	16	US-10-437-963-180124	Sequence 180124,
17	91	5.5	201	15	US-10-369-493-13717	Sequence 13717, A
18	91	5.5	455	15	US-10-425-114-42633	Sequence 42633, A
19	91	5.5	805	15	US-10-424-599-191796	Sequence 191796,
20	91	5.5	1319	16	US-10-437-963-197783	Sequence 197783,
21	91	5.5	1788	16	US-10-437-963-197780	Sequence 197780,
22	90.5	5.5	419	15	US-10-424-599-260387	Sequence 260387,
23	90.5	5.5	578	15	US-10-418-861B-32	Sequence 32, Appl
24	90.5	5.5	661	15	US-10-369-493-2558	Sequence 2558, Ap
25	90.5	5.5	716	15	US-10-369-493-2175	Sequence 2175, Ap
26	90	5.4	559	9	US-09-364-847-21	Sequence 21, Appl
27	90	5.4	856	9	US-09-364-847-33	Sequence 33, Appl
28	90	5.4	856	9	US-09-364-847-35	Sequence 35, Appl
29	90	5.4	1500	16	US-10-437-963-111265	Sequence 111265,
30	90	5.4	5215	9	US-09-861-289-2	Sequence 2, Appl1
31	90	5.4	5215	9	US-09-860-846-2	Sequence 2, Appl1
32	90	5.4	5215	10	US-09-988-384B-2	Sequence 2, Appl1
33	90	5.4	5215	10	US-09-836-821-2	Sequence 2, Appl1
34	90	5.4	5215	14	US-10-271-889-45	Sequence 45, Appl
35	89.5	5.4	203	15	US-10-369-493-8604	Sequence 8604, Ap
36	89	5.4	407	15	US-10-424-599-267113	Sequence 267113,
37	89	5.4	1056	15	US-10-282-174-472	Sequence 472, App
38	89	5.4	1056	15	US-10-282-174-474	Sequence 474, App
39	89	5.4	1056	15	US-10-282-174-476	Sequence 476, App
40	89	5.4	1056	17	US-10-600-009-472	Sequence 472, App
41	89	5.4	1056	17	US-10-600-009-476	Sequence 476, App
42	89	5.4	1056	17	US-10-600-009-476	Sequence 476, App
43	89	5.4	1057	16	US-10-714-796-4	Sequence 4, Appl1
44	88.5	5.3	656	15	US-10-369-493-18559	Sequence 18559, A
45	88.5	5.3	799	16	US-10-437-963-112603	Sequence 112603,

## ALIGNMENTS

```
RESULT 1
US-10-463-676-4
; Sequence 4, Application US/10463676
; Publication No. US20040175375A1
; GENERAL INFORMATION:
; APPLICANT: Park, Bum-Joon
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method
; FILE REFERENCE: 012679-091
; CURRENT APPLICATION NUMBER: US/10/463,676
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: KR 10-2003-13058
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(312)
; OTHER INFORMATION: 1-312 amino acid sequence of p38/JTV-1
US-10-463-676-4
```

```
Query Match      86.8%; Score 1436; DB 16; Length 312;
Best Local Similarity 87.5%; Pred. No. 2.6e-137;
Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEMYGVKPYHGASAPLRYELPTCMYRLPVVHSKTTSPATDAGVQETSPPSLQALESSROD 60
DB 1 MEMYGVKPYHGASAPLRYELPTCMYRLPVVHSKTTSPATDAGVQETSPPSLQALESSROD 60
QY 61 DILKRLYEIKAAVDGISKVHITPDADLDVNTILQADEPTLATNTLNLNSVLGKDYGALK 120
|||||
```

```
Db 61 DILKRLYELKAAVDSKMIQTDPADLDVTNIIQADEPTTLITNADLNSVYKDYGALK 120
Qy 121 DIVINANPASPPLSLVTLHRLCEHRRVYSTVHTSSVKNVPENLVKCGEQARKQSRHE 180
Db 121 DIVINANPASPPLSLVTLHRLCEHRRVYSTVHTSSVKNVPENLVKCGEQARKQSRHE 180
Qy 181 YQGLTLMKNVKTQMKFSVQTMCPIDEGGNARLFSLPQKNAVAVTLLIDSWVIA 240
Db 181 YQGLTLMKNVKTQMKFSVQTMCPIDEGGNARLFSLPQKNAVAVTLLIDSWVIA 240
Qy 241 MPOLRGSSSKKAAVRSNNSALGRSPMLVGNELTVADVVLMSVLTQOTGSSGAPITVQ 300
Db 241 IFOLKGGSSKKAAVFRSMNSALGKSPMLAGNELTVADVVLMSVLTQOTGSSGAPITVQ 300
Qy 301 RMLKSCENLAPF 312
Db 301 RWRBSCENLAPF 312
```

```
RESULT 2
US-10-463-676-6
; Sequence 6; Application US/10463676
; Publication No. US20040175375A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Sunghoon
; APPLICANT: Park, Bum-Joon
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method
; FILE REFERENCE: 012679-091
; CURRENT APPLICATION NUMBER: US/10/463,676
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: KR 10-2003-13058
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(229)
; OTHER INFORMATION: 84-312 amino acid sequence p38/JTV-1
US-10-463-676-6
```

```
Query Match 64.6%; Score 1069; DB 16; Length 229;
Best Local Similarity 88.2%; Pred. No. 4e-100;
Matches 202; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

Qy 84 DADLDVTNIIQADEPTTLATNTLIDNSVYKDYGALKDIVINANPASPPLSLVTLHRLC 143
Db 1 DADLDVTNIIQADEPTTLITNADLNSVYKDYGALKDIVINANPASPPLSLVTLHRLC 60
Qy 144 EHRVYSTVHTSSVKNVPENLVKCGEQARKQSRHEYQGLTLMKNVKTQMKFSVQ 203
Db 61 EHRVYSTVHTSSVKNVPENLVKCGEQARKQSRHEYQGLTLMKNVKTQMKFSVQ 120
Qy 204 MCPIDEGGNARLFSLPQKNAVAVTLLIDSWVIA MPOLRGSSSKKAAVFRSMNSAL 263
Db 121 MCPIDEGGNARLFSLPQKNAVAVTLLIDSWVIA IFOLKGGSSKKAAVFRSMNSAL 180
Qy 264 GRSPMLVGNELTVADVVLMSVLTQOTGSSGAPITVQ RMLKSCENLAPF 312
Db 181 GRSPMLVGNELTVADVVLMSVLTQOTGSSGAPITVQ RWRBSCENLAPF 229
```

```
RESULT 3
US-10-463-676-5
; Sequence 5; Application US/10463676
; Publication No. US20040175375A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Sunghoon
; APPLICANT: Park, Bum-Joon
```

```
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method
; FILE REFERENCE: 012679-091
; CURRENT APPLICATION NUMBER: US/10/463,676
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: KR 10-2003-13058
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(161)
; OTHER INFORMATION: 1-161 amino acid sequence of p38/JTV-1
US-10-463-676-5
```

```
Query Match 44.4%; Score 735; DB 16; Length 161;
Best Local Similarity 88.8%; Pred. No. 2.5e-66;
Matches 143; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MEMVQKHYHGGASPLRVELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSIQALSRQD 60
Db 1 MEMVQKHYHGGASPLRVELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSIQALSRQD 60
Qy 61 DILKRLYELKAAVDSKMIQTDPADLDVTNIIQADEPTTLITNADLNSVYKDYGALK 120
Db 61 DILKRLYELKAAVDSKMIQTDPADLDVTNIIQADEPTTLITNADLNSVYKDYGALK 120
Qy 121 DIVINANPASPPLSLVTLHRLCEHRRVYSTVHTSSVKNV 161
Db 121 DIVINANPASPPLSLVTLHRLCEHRRVYSTVHTSSVKNV 161
```

```
RESULT 4
US-10-437-963-126132
; Sequence 126132; Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalick, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; PRIOR FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 126132
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Oryza sativa
; OTHER INFORMATION: Clone ID: PAT_MFT4530_28709C.1.pcp
US-10-437-963-126132
```

```
Query Match 6.4%; Score 105.5; DB 16; Length 925;
Best Local Similarity 20.3%; Pred. No. 0.51;
Matches 64; Conservative 46; Mismatches 107; Indels 99; Gaps 13;

Qy 25 YLPRVHAKTSPADDAHVQETSPSLQALSRQDI--LKLRLYELKAAVDSKMIHT 82
Db 152 YLWDNIIKVTGPV-----TIDPRLOALYKKTTELVGIEKSEKLVILSLGDVHA 203
Qy 83 PDADLDVTNIIQAD--EPTTLATNTLIDNSVYKDYGA-----LQDIVINAN 127
Db 83 PDADLDVTNIIQAD--EPTTLATNTLIDNSVYKDYGA-----LQDIVINAN 127
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 23, 2005, 13:34:50 ; Search time 13.3213 Seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-5

Perfect score: 1655

Sequence: 1 MMYGVKHPYHGSAPLRLVEL.....RWLKSCENLAPRYTALQLLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126.5	7.6	719	2	T52043
2	112	6.8	728	2	T01200
3	104	6.3	980	2	T30089
4	99	6.0	691	2	B89797
5	98.5	6.0	222	2	B82441
6	98.5	6.0	377	2	P69172
7	98	5.9	690	2	A24545
8	96.5	5.8	1440	1	SYHOQT
9	95.5	5.8	753	2	T24869
10	95	5.7	488	1	H64313
11	95	5.7	552	2	D82421
12	95	5.7	2655	2	D96595
13	94.5	5.7	2471	2	T42877
14	94	5.7	435	2	T01567
15	94	5.7	436	2	C86719
16	93.5	5.6	1064	2	B86465
17	93	5.6	490	2	D84999
18	92	5.6	754	2	S37403
19	91.5	5.5	436	2	T51237
20	91	5.5	1265	2	T47626
21	90.5	5.5	359	2	T37921
22	90.5	5.5	456	2	C84764
23	90.5	5.5	642	2	T39490
24	90.5	5.5	716	2	T37630
25	90	5.4	437	2	T51238
26	90	5.4	559	2	A38604
27	89.5	5.4	2954	2	T14156
28	89.5	5.4	317	2	T05528
29	89.5	5.4	1597	2	S65053

30	89.5	5.4	1714	1	S18644	multifunctional am
31	89.5	5.4	4976	2	T14165	peptide synthetase
32	89	5.4	184	2	S74356	glutathione S-tran
33	89	5.4	763	2	AE2443	penicillin-binding
34	89	5.4	1023	2	T31669	neutral zinc finger
35	89	5.4	1056	1	G02157	kinesin-like spind
36	89	5.4	1247	2	A33812	interphotoreceptor
37	88.5	5.3	602	2	AB3542	gtp-binding protei
38	88.5	5.3	656	2	H84206	acetyl-CoA synthet
39	88	5.3	821	2	T24728	hypothetical prote
40	88	5.3	2279	2	T42531	acetyl-CoA carboxy
41	88	5.3	2280	2	T38906	acetyl-CoA carboxy
42	87.5	5.3	474	2	B70597	hypothetical prote
43	87.5	5.3	820	2	H82302	ATP-dependent heli
44	87	5.3	570	2	T15763	hypothetical prote
45	86	5.2	906	2	T45158	pre-mRNA splicing

#### ALIGNMENTS

##### RESULT 1

T52043 Probable glutamate-tRNA ligase (BC 6.1.1.17) [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #ext\_change 09-Jul-2004

C/Accession: T52043

R/Day, I.S.; Golovkin, M.; Reddy, A.S.

Biochim. Biophys. Acta 1399, 219-24, 1998

A/Title: Cloning of the cDNA for glutamyl-tRNA synthetase from Arabidopsis thaliana.

A/Reference number: Z24836; MUID:9765600; PMID:9765600

A/Accession: T52043

A/Status: preliminary; translated from GB/EMBL/DBDJ

A/Molecule type: mRNA

A/Residues: 1-719 <DAY>

A/Cross-references: UNIPROT:082462; EMBL:AF067773; PTDN:AAC36469.1

A/Suprafamily: Yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology

C/Keywords: ligase

Query Match 7.6%; Score 126.5; DB 2; Length 719;

Best Local Similarity 28.0%; Pred. No. 0.017;

Matches 49; Conservative 22; Mismatches 57; Indels 47; Gaps 8;

QY 128 PASPPSLVLRILLCERVLSTVTHSSVKNVPENLVKCFGEQARKSRHEVQGFLL 187

DB 10 PESPPLSVLVALSLASPV-----TIDSSAAATTVSEFV-----SDGRKLN-----GATV 55

QY 188 IKNVVPKTMKFSVQTMCPIDEGGNIAFLFSLFGQKHNAVTLTLIDSWVDIAMPOLREG 247

DB 56 LARVY-----GREAKKLPDFYG--NNAFPSSQIDENVVDVASY--FSSG 94

QY 248 SSEKKAIVFRSNMNSALGRSPMTVGNELTVADVVLVSLQOTGGSGAAPTNQWR 302

DB 95 SEFENAC--GRVDKYLSESTFLVGHSLSTADVAIWALAGTG-----QRW 137

##### RESULT 2

T01200 Probable glutamate-tRNA ligase (BC 6.1.1.17) F21E10.12 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #ext\_change 09-Jul-2004

C/Accession: T01200

R/Davieson, S.; Rohlfing, T.; David, M.; O'Brian, D.

submitted to the EMBL Data Library, April 1998

A/Description: The sequence of A. thaliana F21E10.

A/Reference number: Z14258

A/Accession: T01200

A/Status: translated from GB/EMBL/DBDJ

A/Molecule type: DNA

A/Residues: 1-728 <DAY>

A/Cross-references: UNIPROT:065253; EMBL:AF058914; NID:93047074; PID:93047084; GSPDB:GNC

A/Experimental source: cultivar Columbia

C/Genetics:

A:Gene: ATSP.F21B10.12  
A:Map position: 5  
A:Introns: 47/2; 89/3; 141/1; 503/3; 659/3  
C:Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
F:223-499/Domain: glutamine-tRNA ligase homology <EG>

Query Match 6.8%; Score 112; DB 2; Length 728;  
Best Local Similarity 26.6%; Pred. No. 0.29;  
Matches 49; Conservative 22; Mismatches 57; Indels 56; Gaps 9;

QY 128 PASPPLSLVHRLLCERYRLVSTHTHSVKNVPEINLVKCGEQARKQSHHEVQLGFTL 187  
DB 10 PESPPLSVVALSLASPV---TIDSSAAATTVSPFV---SDGRKLN-----GATV 55  
QY 188 IMKNVPKQMKFSVOTMCPIDEGENIARPLPSLFGQKINAVTLT-----LIDSWVD 238  
DB 56 LIRYV-----GRSAKKLPDPYG--NNAPDSQVSILCIMKIDEMVD 95  
QY 239 IAMFOLREGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVVMVYQQTGSSGGAAPT 298  
DB 96 YASV-FSSGSEFENMC--GRVDKILSSTFLVHSLTADVAIVSALAGTG----- 143  
QY 239 VQRW 302  
DB 144 -QRW 146

RESULT 3  
T30089  
Probable zinc proteinase (EC 3.4.24.-) CO2G6.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30089  
R:Bentley, D.; Kemp, K.; Scheer, P.  
A:Description: The sequence of *C. elegans* cosmid CO2G6.  
A:Reference number: Z20734  
A:Accession: T30089  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-980 <BEN>  
A:Cross-references: UNIPROT:Q17592; EMBL:U55372; PDB:AAA9001.1; GSPDB:GN00023; CESP:CO  
A:Experimental source: strain Bristol N2; clone CO2G6  
C:Genetics:  
A:Gene: CESP.CO2G6.1  
A:Map position: 5  
A:Introns: 25/3; 266/3; 540/3; 585/3; 786/1; 898/2  
C:Superfamily: insulin-degrading enzyme (IDE)  
C:Keywords: hydrolase; metalloproteinase; zinc  
F:70/74/Binding site: zinc (His) #status predicted  
F:73/Active site: Glu #status predicted

Query Match 6.3%; Score 104; DB 2; Length 980;  
Best Local Similarity 20.1%; Pred. No. 2.1;  
Matches 86; Conservative 67; Mismatches 133; Indels 142; Gaps 22;

QY 10 HGSASRLRELPCTMRPLNVHASKTTPATDAGHOVETSEPSLQALSRQDILKRLYE- 68  
DB 276 HKPGSLVLEIKRLGW-VNSLSDSNITIAAGFIINVTMDLSTGLLEN-VDEIIQLMLNY 333  
QY 69 ---LKAADVGLSKMHTPPADL-DVT-NILQADEPTTLATNT-----LDLANSVLGADY- 116  
DB 334 IGMKLSF--GPGQWHDLELADLDVYKFKQKQMKKAINIASLQYTFEHLISRYL 391  
QY 117 -----GALKDVIIVANPASPLSLVLRLLCERY----- 147  
DB 392 LTKYEPERIKELSLTLPSTN-----MLVYVVSQKFKQEGQNTNPEVYGTETMKTDISPE 445  
QY 148 -----VLSTVH---THSSVKNVPEINLVKCGEQARKQSHHEVQL-----GFTLIW----- 189  
DB 446 KMKKYNALTKTSHALHLPKKN--EYIATVFGQKQKRESVKNKHPKLLSDGWSKVFQKOD 503

QY 190 --KNVPKQMKFSVOT-----MCPLEG-----GNIAPLPSL 220  
DB 504 DENMKQKQKFPALTPPIYSQNPRLISLSLWMCFCDDILSEETYNALAGCCQFELSP 563  
QY 221 FG-----QKKNVAVTL-----TLIDSWVDIAMFOLREGSSKE 251  
DB 564 FGQKQKSTQDGERAEHRASTLHVYGYDEKQPLFVHNLISCMINFKIDRTREPEVLFSJK- 622  
QY 252 KAAPFRSMNSALGRSPMLVGNELTVADVVMVYQQTGSSGGAAPTNAVQRMKSCEN 308  
DB 623 -----RLTNHAPSGQVLLTQHYNQLLIVDKV-WSEKQDLAVCDSVLTLENVQGFAR--EM 674  
QY 309 LAPFSTAL 316  
DB 675 LQAFHML 682

RESULT 4  
B89797  
Glycerol ester hydrolase [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: B89797  
R:Kuroda, M.; Ohta, T.; Uchiyama, T.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shida, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: B89797  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-691 <KUR>  
A:Cross-references: UNIPROT:Q09WQ6; GB:BA000018; PID:G13700235; PDB:BAB41533.1; GSPDB:G  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: geh  
C:Superfamily: *Staphylococcus triacylglycerol* lipase

Query Match 6.0%; Score 99; DB 2; Length 691;  
Best Local Similarity 20.2%; Pred. No. 3.4;  
Matches 65; Conservative 53; Mismatches 110; Indels 94; Gaps 15;

QY 1 MPWQVPEYVGGAPRLRELPCTMRPLNVHASKTTPATD-----GHVETSEPSL 52  
DB 172 IPQKVLPHNDKAAFPSTTPS-----NDTAPKSTKADPATTKHPNOODTHQPAH 223  
QY 53 QALSRQDILKRLYEIKAAVDGLSKMI---HTPDADL-----VTNIIQADEPTTL 101  
DB 224 QIIDAKQDDIV-RQSEKQPVGDLSGHIGQNSPEKPTDKNTDKQLIDALQDA--PKTR 280  
QY 102 AYTTLIDNSVLGQDVYQALKDVIIVANPASPLSLVLRLLCERYVSTVTHSSVKNV 161  
DB 281 STTNAAAD-----AKVRLPKANQVQV-----LNKKPV--VFVHGFGLV 318  
QY 162 PENLVKCFG-----EQAARKQSHHEVQLGFTLIWKNVPK-TQMKFSQYTCPIEG 209  
DB 319 GDAAPALVRYWQNGKFKYIEELRKQGVNVHQAASVAFSGSNVDRVAVELYIYIKGRVDYG 378  
QY 210 EGNIAPLPSLFGQKINAV-----TLTIDSWVDIAMFOLREGSSKE 249  
DB 379 AAIDAKYGHERYGKYTKGIMPMWEPKQVNLVGHSMGGQTRLMEEF-----LRNG-N 430  
QY 250 KEKAAPFRSMNSALGRSPMLVG 271  
DB 431 KEELIAYHKAAGSEI--SPLFTG 450

RESULT 5  
G82441  
Probable glutathione S-transferase VCA0584 [imported] - *Vibrio cholerae* (strain N16961)  
C:Species: *Vibrio cholerae*  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:33:04 ; Search time 60.8145 Seconds  
(without alignments)  
2694.512 Million cell updates/sec

Title: US-10-622-817-5

Perfect score: 1655

Sequence: 1 MPWGVKPYHGSAPLRLVEL.....RLMKSCENLAPSTALQLLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1655	100.0	320	Q8R010	Q8R010 mus musculus
2	1652	99.8	320	Q8RZY6	Q8RZY6 mus musculus
3	1573	95.0	320	1 MCA2_CRIGR	Q9WY77 cricetus
4	1464	88.5	320	1 MCA2_HUMAN	Q13155 homo sapien
5	1460	88.2	320	2 Q96CZ5	Q96CZ5 mus musculus
6	1412	85.3	280	2 Q8R3V2	Q8R3V2 mus musculus
7	1153.5	69.7	311	2 Q6DK86	Q6DK86 xenopus tro
8	1146.5	69.3	311	2 Q6IND4	Q6IND4 xenopus lae
9	1146.5	69.3	311	2 Q7ZYD7	Q7ZYD7 xenopus lae
10	926.5	56.0	321	2 Q7J3C0	Q7J3C0 brachydanio
11	247	14.9	340	2 Q7Q7A3	Q7Q7A3 anopheles g
12	224	13.5	301	2 Q7KUM5	Q7KUM5 drosophila
13	224	13.5	313	2 Q8T060	Q8T060 drosophila
14	224	13.5	322	2 Q6NKM4	Q6NKM4 drosophila
15	216.5	13.1	334	1 MCA2_DROME	Q9VUR3 drosophila
16	126.5	7.6	719	2 Q82462	Q82462 arabidopsis
17	112	6.8	728	2 Q6S253	Q6S253 arabidopsis
18	107	6.5	715	2 Q91LZ8	Q91LZ8 cryza sativ
19	106	6.4	328	2 Q86X73	Q86X73 homo sapien
20	106	6.4	559	2 Q8RPZ6	Q8RPZ6 pseudomona
21	106	6.4	869	2 Q6P57	Q6P57 homo sapien
22	106	6.4	1721	2 Q7NID9	Q7NID9 glisobacter
23	104	6.3	980	2 Q17592	Q17592 caenorhabdi
24	102.5	6.2	420	2 Q8PQ00	Q8PQ00 coerynebacte
25	102	6.2	559	2 Q939A8	Q939A8 pseudomona
26	102	6.2	1048	2 Q7N3E2	Q7N3E2 photorhabdu
27	101.5	6.1	715	2 Q91B77	Q91B77 white spot
28	101.5	6.1	730	2 Q91BF5	Q91BF5 white spot
29	101.5	6.1	913	2 Q6DRB3	Q6DRB3 brachydanio
30	101.5	6.1	913	2 Q6PQ04	Q6PQ04 brachydanio
31	100.5	6.1	555	2 Q7MEV7	Q7MEV7 vibrrio vuln

32	100.5	6.1	555	2	Q8D7U3	Q8D7U3 vibrrio vuln
33	100.5	6.1	982	2	Q8K2J4	Q8K2J4 mus musculus
34	99.5	6.0	922	2	Q6TEM5	Q6TEM5 brachydanio
35	99	6.0	559	2	Q8R067	Q8R067 pseudomona
36	99	6.0	690	1	LIP_STRAW	Q8Y62 staphylococ
37	99	6.0	690	2	Q6GCF1	Q6GCF1 staphylococ
38	99	6.0	691	2	Q79827	Q79827 staphylococ
39	99	6.0	691	2	Q99WQ6	Q99WQ6 staphylococ
40	99	6.0	691	2	Q7A7P2	Q7A7P2 staphylococ
41	98.5	6.0	222	2	Q9KWM5	Q9KWM5 vibrrio chol
42	98.5	6.0	377	2	Q26161	Q26161 methanobact
43	98.5	6.0	1597	2	Q8V9A2	Q8V9A2 crucifer to
44	98	5.9	240	2	Q8BTR1	Q8BTR1 mus musculus
45	98	5.9	690	1	LIP_STRAW	P10335 staphylococ

## ALIGNMENTS

RESULT 1  
Q8R010  
ID Q8R010 PRELIMINARY; PRT; 320 AA.  
AC Q8R010;  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Jtv1-pending protein.  
GN Name=Jtv1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N, and FVB/N-3; TISSUE=Mammary tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,  
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny K.C., Harte S., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley A.C., Touchman J.W., Green E.D., Dickson M.C.,  
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Maier M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N, TISSUE=Mammary tumor. C3;  
RA Strausberg R.;  
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N-3; TISSUE=Mammary tumor;  
RA Strausberg R.;  
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL, BC024410; AAH24410.1; -  
DR EMBL, BC026972; AAH26972.1; -  
DR MGD, MGI:2385237; Jtv1.  
DR InterPro; IPR004046; GST\_Cterm.  
DR InterPro; IPR010987; GST\_C\_1like.  
DR Pfam; PF00043; GST\_C\_1.  
DR SEQUENCE 320 AA; 35396 MW; 1745D7E4BC3670D CRC64;

Query Match 100.0%; Score 1655; DB 2; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 2,6e-128;  
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYQKPYHGSAPRLVELPTCMYRLPNVHSKTTPATDAGHVEETSEPSLQALSHOD 60  
 DB 1 MPWYQKPYHGSAPRLVELPTCMYRLPNVHSKTTPATDAGHVEETSEPSLQALSHOD 60  
 QY 61 DILKRLYEKAADVGSKMIHTPPDADLVNTNIIQADEPTTLATNTLDNSVLGKDYALK 120  
 DB 61 DILKRLYEKAADVGSKMIHTPPDADLVNTNIIQADEPTTLATNTLDNSVLGKDYALK 120  
 QY 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 DB 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 QY 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 DB 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 QY 181 YOLGFTLLWKVNPKTQMKFSVQTMCPLEGEGNIARFLPSLFGQKHNAVTLTLIDSWDIA 240  
 DB 181 YOLGFTLLWKVNPKTQMKFSVQTMCPLEGEGNIARFLPSLFGQKHNAVTLTLIDSWDIA 240  
 QY 241 MFOLREGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTNVQ 300  
 DB 241 MFOLREGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTNVQ 300  
 QY 301 RWLKSCEMLAPFTALQLLK 320  
 DB 301 RWLKSCEMLAPFTALQLLK 320

RESULT 2  
 QGR2Y6 PRELIMINARY; PRT; 320 AA.  
 AC QGR2Y6;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Jty1-Pending protein.  
 GN Name=Jty1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Heiton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RA Strausberg R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, BC026958; AAH26958.1; -.

DR MGI:2385237; Jty1.  
 DR InterPro: IPR004046; GST\_Cterm.  
 DR InterPro: IPR010987; GST\_C1like.  
 DR Pfam: PF00043; GST\_C; 1.  
 SQ SEQUENCE 320 AA; 35423 MW; 1C21FLA74C9882B4 CRC64;

Query Match 99.8%; Score 1652; DB 2; Length 320;  
 Best Local Similarity 99.7%; Pred. No. 4,6e-128;  
 Matches 319; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYQKPYHGSAPRLVELPTCMYRLPNVHSKTTPATDAGHVEETSEPSLQALSHOD 60  
 DB 1 MPWYQKPYHGSAPRLVELPTCMYRLPNVHSKTTPATDAGHVEETSEPSLQALSHOD 60  
 QY 61 DILKRLYEKAADVGSKMIHTPPDADLVNTNIIQADEPTTLATNTLDNSVLGKDYALK 120  
 DB 61 DILKRLYEKAADVGSKMIHTPPDADLVNTNIIQADEPTTLATNTLDNSVLGKDYALK 120  
 QY 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 DB 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 QY 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 DB 121 DIVINANPASPLSLVLRLLCERYRLSTVHTSHSVKNVPELVKCFGEQARKQRHE 180  
 QY 181 YOLGFTLLWKVNPKTQMKFSVQTMCPLEGEGNIARFLPSLFGQKHNAVTLTLIDSWDIA 240  
 DB 181 YOLGFTLLWKVNPKTQMKFSVQTMCPLEGEGNIARFLPSLFGQKHNAVTLTLIDSWDIA 240  
 QY 241 MFOLREGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTNVQ 300  
 DB 241 MFOLREGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTNVQ 300  
 QY 301 RWLKSCEMLAPFTALQLLK 320  
 DB 301 RWLKSCEMLAPFTALQLLK 320

RESULT 3  
 MCA2\_CRIGR STANDARD; PRT; 320 AA.  
 AC 09WVW7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Multisynthetase complex auxiliary component p38.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 NC NCB1\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=9906915; PubMed=9878398; DOI=10.1006/jmbi.1998.2316;  
 RA Quevillon S., Robinson J.-C., Berthomieu E., Slatkova M., Mirande M.,  
 RT "Molecular assembly of aminoacyl-tRNA synthetases:  
 identification of protein-protein interactions and characterization of  
 a core protein.";  
 RL J. Mol. Biol. 285:183-195(1999).  
 RT -1- FUNCTION: Probable core protein of the multisynthetase complex  
 that serves as a template for the assembly of the supramolecular  
 structure.  
 CC -1- SUBUNIT: Component of the multisynthetase complex which is  
 composed of a bifunctional glutamyl-prolyl-tRNA synthetase, the  
 monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
 arginyl, and aspartyl-tRNA synthetases as well as three auxiliary  
 proteins, p18, p48 and p43.  
 CC -1- SIMILARITY: Contains 1 GST-like domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>)

or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch).

CC EMBL; AF072727; AAD38422.1; -;  
 CC InterPro; IPR010987; GST\_C\_like.  
 DR InterPro; IPR004046; GST\_C-term.  
 DR Pfam; PF00043; GST\_C; 1.  
 KW Protein biosynthesis.  
 SQ SEQUENCE 320 AA; 35433 MW; 6D24E03ABEC810A CRC64;

Query Match 95.0%; Score 1573; DB 1; Length 320;  
 Best Local Similarity 94.4%; Pred. No. 1.6e-121;  
 Matches 302; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 MEMYQVKPYHGSGAPRLVRLPTCMYRLPNVHSKTTSPATDAGHVQETSEPSILOAESRD 60  
 DB 1 MEMYQVKSHYGSGAPRLVRLPTCMYRLPNVHSKTTSPADAGHVQETSEPSILOAESRD 60  
 QY 61 DILKRLYLKAAVDSKMIHPDADLVYNTIIQADEPTTLATNTLDLSVLEKDYGALK 120  
 DB 61 DILKRLYLKAAVDSKMIHPDADLVYNTIIQADEPTTLATNTLDLSVLEKDYGALK 120  
 QY 121 DIVINANPASPPLSLVLRLLCERRVLTSTVHTSSVKNVPELVKCTGEQARKQSRHE 180  
 DB 121 DIVINANPASPPLSLVLRLLCERRVLTSTVHTSSVKNVPELVKCTGEQARKQSRHE 180  
 QY 181 YOLGFTLWKQVPTKQMKFSVOTMCPIDEGGNARFLFSLFGQKHAIVTLTLDLSVNDIA 240  
 DB 181 YOLGFTLWKQVPTKQMKFSVOTMCPIDEGGNARFLFSLFGQKHAIVTLTLDLSVNDIA 240  
 QY 241 MFOLREGSGSKKAAVFRSMNSALGRSPMLVGNELTVADVVLVSLQOTGGSGGAPLTNYQ 300  
 DB 241 MFOLREGSGSKKAAVFRSMNSALGRSPMLVGNELTVADVVLVSLQOTGGSGGAPLTNYQ 300  
 QY 301 RMLKSCENLAPSTALQLK 320  
 DB 301 RMLKSCENLAPSTALQLK 320

#### RESULT 4

MCAL2\_HUMAN STANDARD; PRT; 320 AA.  
 ID Q13155; Q9PIL2;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Multisynthetase complex auxiliary component p38 (JTV-1 protein)  
 DE (PRO00992).  
 GN Name=JTV1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCB1\_TaxID=9606;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96115582; PubMed=8666379;  
 RA Nicolson N.C., Kinzler K.W., Vogelstein B.;  
 RT "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts  
 and a novel overlapping gene.";  
 RL Genomics 29:329-334 (1995).  
 RT 12;  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=lymph;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.S., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abirmson R.D., Miliady S.J.,  
 RA Boeck S.A., Mcwan P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,  
 RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butlerfield Y.S.N., Krzywicki M.I., Skalska U., Shailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN (3)  
 RP SEQUENCE OF 197-320 FROM N.A.  
 RC TISSUE=Fetal liver;

RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
 RA Liu M., He F.;  
 RT "Functional prediction of the coding sequences of 121 new genes  
 RT deduced by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBA databases.  
 RN (4)

RP INTERACTION WITH FUBP1.

RX MEDLINE=22716800; PubMed=12819782; DOI=10.1038/ng1182;  
 RA Kim M.J., Park B.-J., Kang Y.-S., Kim H.J., Park J.-H., Kang J.W.,  
 RA Lee S.W., Han J.M., Lee H.-W., Kim S.;

RT "Downregulation of FUSE-binding protein and c-myc by cRNA synthetase  
 RT cofactor p38 is required for lung cell differentiation.";  
 RL Nat. Genet. 34:330-336 (2003).

CC -1- FUNCTION: Probable core protein of the multisynthetase complex  
 CC that serves as a template for the assembly of the supramolecular  
 CC structure. Mediates ubiquitination of FUBP1 and its degradation by  
 CC the proteasome.

CC -1- SUBUNIT: Component of the multisynthetase complex which is  
 CC comprised of a bifunctional glutamyl-prolyl-cRNA synthetase, the  
 CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
 CC arginyl, and aspartyl-cRNA synthetases as well as three auxiliary  
 CC proteins, p18, p48 and p43. Binds FUBP1.

CC -1- SIMILARITY: Contains 1 GST-like domain.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
 CC frameshift in position 312.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

CC EMBL; U24169; AAC50391.1; ALT FRAME.

DR EMBL; BC002853; AAH02853.1; -;

DR EMBL; BC010156; AAH010156.1; -;

DR EMBL; AF116151; AAF71039.1; -;

DR H-InvDB; HIX0006460; -;

DR MIM; 600859; -;

DR InterPro; IPR010987; GST\_C\_like.

DR InterPro; IPR004046; GST\_C-term.

DR Pfam; PF00043; GST\_C; 1.

KW Protein biosynthesis.

KW SEQUENCE 320 AA; 35349 MW; F253726B63C12BAB CRC64;

Query Match 88.5%; Score 1464; DB 1; Length 320;  
 Best Local Similarity 87.2%; Pred. No. 1.6e-112;  
 Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEMYQVKPYHGSGAPRLVRLPTCMYRLPNVHSKTTSPATDAGHVQETSEPSILOAESRD 60  
 DB 1 MEMYQVKSHYGSGAPRLVRLPTCMYRLPNVHSKTTSPADAGHVQETSEPSILOAESRD 60  
 QY 61 DILKRLYLKAAVDSKMIHPDADLVYNTIIQADEPTTLATNTLDLSVLEKDYGALK 120  
 DB 61 DILKRLYLKAAVDSKMIHPDADLVYNTIIQADEPTTLATNTLDLSVLEKDYGALK 120  
 QY 121 DIVINANPASPPLSLVLRLLCERRVLTSTVHTSSVKNVPELVKCTGEQARKQSRHE 180  
 DB 121 DIVINANPASPPLSLVLRLLCERRVLTSTVHTSSVKNVPELVKCTGEQARKQSRHE 180

```
Db 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPEENILKCFGEONKQOPROD 180
Qy 181 YOLGFTLLIKWNPVKTKOMKRSVOTMCPPIEGEINIAFLFSLFGQKNAVTLTLIDSWDIA 240
Db 181 YOLGFTLLIKWNPVKTKOMKRSVOTMCPPIEGEINIAFLFSLFGQKNAVTLTLIDSWDIA 240
Qy 241 MFOLREGSSKEKAARFRSNNSALGRSPWLVGNELTVDVAVYMSVYLOQTGGSSGAAPTNVQ 300
Db 241 IFOLKREGSSKEKAARFRSNNSALGRSPWLAGNELTVADVIMSVYLOQTGGCSVTVPANVQ 300
Qy 301 RWLKSCENIAPFSTALQILK 320
Db 301 RWRMSCENIAPFNTALKILK 320

RESULT 5
096CZ5 PRELIMINARY; PRT; 320 AA.
AC 096CZ5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE JTV1.
GN Name=JTV1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC013630; AAH1330.1; -
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR010987; GST_C_1like.
DR Pfam: PF00043; GST_C_1.
SQ SEQUENCE 320 AA; 35335 MW; 19F14BF758612E08 CRC64;

Query Match 88.2%; Score 1460; DB 2; Length 320;
Best Local Similarity 86.9%; Pred. No. 3,4e-112;
Matches 278; Conservative 16; Mismatches 26; Indels 0; Gaps 0;
```

```
Db 61 DILKRLYEKAADVGLSKMIQTDPADLDVTNIIQADEPTTLTNALDINSVLGKDYALK 120
Qy 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPEENILKCFGEONKQOPROD 180
Db 121 DIVINANPASPLSLVLRLLCEHFRVLTSTVHTSSVSKVPEENILKCFGEONKQOPROD 180
Qy 181 YOLGFTLLIKWNPVKTKOMKRSVOTMCPPIEGEINIAFLFSLFGQKNAVTLTLIDSWDIA 240
Db 181 YOLGFTLLIKWNPVKTKOMKRSVOTMCPPIEGEINIAFLFSLFGQKNAVTLTLIDSWDIA 240
Qy 241 MFOLREGSSKEKAARFRSNNSALGRSPWLVGNELTVDVAVYMSVYLOQTGGSSGAAPTNVQ 300
Db 241 IFOLKREGSSKEKAARFRSNNSALGRSPWLAGNELTVADVIMSVYLOQTGGCSVTVPANVQ 300
Qy 301 RWLKSCENIAPFSTALQILK 320
Db 301 RWRMSCENIAPFNTALKILK 320

RESULT 6
09R3V2 PRELIMINARY; PRT; 280 AA.
ID 09R3V2;
AC 09R3V2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE JTV1.
GN Name=JTV1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC024480; AAH24480.1; -
DR MGD: MGI:2385237; JTV1.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR010987; GST_C_1like.
DR Pfam: PF00043; GST_C_1.
SQ SEQUENCE 280 AA; 31097 MW; E5EB81498983FD2E CRC64;

Query Match 85.3%; Score 1412; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 2.6e-108;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
DB 5 OETSEPSLOALBESRODILKRLYELKAAVDGSKMHTPADLDVTNIILOADEPTLATN 64
QY 105 TLIDNLVIGKDVGLADIVYNANPASPILSLVHLHLCGRYVLSGVYHHSVKNVPE 164
DB 65 TLIDNLVIGKDVGLADIVYNANPASPILSLVHLHLCGRYVLSGVYHHSVKNVPE 124
QY 165 LVKCFEQAOKOSRHEVQLGFTLIMKNVPTKMFVQTMCPTEGEGNIARFLFSGOK 224
DB 125 LVKCFEQAOKOSRHEVQLGFTLIMKNVPTKMFVQTMCPTEGEGNIARFLFSGOK 184
QY 225 HNAVTLTLIDSVWDIAMFOLREGSSKEKAAPFRSNMSALGRSPMLVGNELTVADVILMSV 284
DB 185 HNAVTLTLIDSVWDIAMFOLREGSSKEKAAPFRSNMSALGRSPMLVGNELTVADVILMSV 244
QY 285 LOOTGSSGAAPTNVGRWLKSCENLAPFSTALQLK 320
DB 245 LOOTGSSGAAPTNVGRWLKSCENLAPFSTALQLK 280

RESULT 7
Q6DK86 PRELIMINARY; PRT; 311 AA.
AC Q6DK86;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE MGCG8221 protein.
GN Name=MGCG8221;
OS Xenopus tropicalis (western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_Taxid=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC074561; AAH74561.1; -
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
SQ SEQUENCE 311 AA, 34480 MW, D98F27F73C466154 CRG64;
```

```
Query Match 69.7%; Score 1153.5; DB 2; Length 311;
Best Local Similarity 67.5%; Pred. No. 6.8e-87;
Matches 216; Conservative 49; Mismatches 46; Indels 9; Gaps 3;
```

```
QY 1 MPMYQVYPYHGSAPLRLVELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSLOALBESROD 60
DB 1 MPMYQVYPYCGGE--IQVDLPTCMYRLPNVHQAVSE-----NLGEADPAIQLBESRVE 52
QY 61 DILKRLYELKAAVDGSKMHTPADLDVTNIILOADEPTLATNTLIDNLVIGKDVGLA 120
DB 53 DILKRLYELKAAVDGSKMHTPADLDVTNIILOADEPTLATNTLIDNLVIGKDVGLA 112
QY 121 DIVYNANPASPILSLVHLHLCGRYVLSGVYHHSVKNVPELVKCFEQAOKOSRHE 180
DB 113 DIVYNANPASPILSLVHLHLCGRYVLSGVYHHSVKNVPELVKCFEQAOKOSRHE 172
QY 181 YQLGFTLIMKNVPTKMFVQTMCPTEGEGNIARFLFSGOKHNAVTLTLIDSVWDIA 240
DB 173 YQLGFTLIMKNVPTKMFVQTMCPTEGEGNIARFLFSGOKHNAVTLTLIDSVWDIA 232
QY 241 MFOLREGSSKEKAAPFRSNMSALGRSPMLVGNELTVADVILMSVLOOTGSSGAAPTNV 300
DB 233 IFOLREGSSKEKAAPFRSNMSALGRSPMLVGNELTVADVILMSVLOOTGSSGAAPTNV 291
QY 301 RMLKSCENLAPFSTALQLK 320
DB 292 RMLKSCENLAPFSTALQLK 311
```

```
RESULT 8
Q6IND4 PRELIMINARY; PRT; 311 AA.
AC Q6IND4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE MGCG80304 protein.
GN Name=MGCG80304;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=223441132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
Initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
```

RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, BC072178; AAH72178.1;  
 DR InterPro: IPR004046; GST\_Cterm.  
 DR InterPro: IPR010987; GST\_C\_1-like.  
 DR Pfam: PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34424 MW; D5E8325C18D8751 CRC64;

Query Match 69.3%; Score 1146.5; DB 2; Length 311;  
 Best Local Similarity 68.1%; Pred. No. 2.6e-86;  
 Matches 218; Conservative 44; Mismatches 49; Indels 9; Gaps 3;

QY 1 MPMTQVKEPHGSGAPLRYELPTCMYRLPNVSKTTPATAGHVOETSEPSLOAESRQD 60  
 DB 1 MPMTQVKEPHGSGA--IQVDLPTCMYRLPNVSKTTPATAGHVOETSEPSLOAESRQD 52  
 QY 61 DILKRIYELKAAVGLSKMHTPPADLDVTNIIQADEPTLATNTLDINSVLGKDYGALK 120  
 DB 53 DILKRIYELKAAVGLSKMHTPPADLDVTNIIQADEPTLATNTLDINSVLGKDYGALK 112  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTSSVKVNPENLVKCFEGQARKOSRHE 180  
 DB 113 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTSSVKVNPENLVKCFEGQARKOSRHE 172  
 QY 181 YOLGFTLIMKNVPTQMKFSVQTMCPLEGSGNIAEFLFSIFGQKHNAAVTLTLDINSV 240  
 DB 173 YOLGFTLIMKNVPTQMKFSVQTMCPLEGSGNIAEFLFSIFGQKHNAAVTLTLDINSV 232  
 QY 241 MFOLREGSSKEKAAVFRSNMISALGRSPMLVGNELTVADVVLMSVLQOTGSSGAAPTNO 300  
 DB 233 IFOLREGSSKEKAAVFRSNMISALGRSPMLVGNELTVADVVLMSVLQOTGSSGAAPTNO 291  
 QY 301 RMLKSCENLAPFSTALQLEK 320  
 DB 292 KMKKSCENLAPFSTALQLEK 311

## RESULT 9

ID Q7ZXD7 PRELIMINARY; PRT; 311 AA.  
 AC Q7ZXD7;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE tlv1-prov protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenopodidae; Xenopus.  
 NC NCB1\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,  
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.C., Harte S., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Krzywinski M.I., Skalska U., Smalios D.B., Schermer A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.,  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT Initiative.";  
 RL Dev. Dyn. 225:384-391 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.,  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, BC043832; AAH43832.1;  
 DR InterPro: IPR004046; GST\_Cterm.  
 DR InterPro: IPR010987; GST\_C\_1-like.  
 DR Pfam: PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34647 MW; 2F08C9DA60DD63BE CRC64;

Query Match 69.3%; Score 1146.5; DB 2; Length 311;  
 Best Local Similarity 68.1%; Pred. No. 2.6e-86;  
 Matches 218; Conservative 44; Mismatches 49; Indels 9; Gaps 4;

QY 1 MPMTQVKEPHGSGAPLRYELPTCMYRLPNVSKTTPATAGHVOETSEPSLOAESRQD 60  
 DB 1 MPMTQVKEPHGSGA--IQVDLPTCMYRLPNVSKTTPATAGHVOETSEPSLOAESRQD 52  
 QY 61 DILKRIYELKAAVGLSKMHTPPADLDVTNIIQADEPTLATNTLDINSVLGKDYGALK 120  
 DB 53 DILKRIYELKAAVGLSKMHTPPADLDVTNIIQADEPTLATNTLDINSVLGKDYGALK 112  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTSSVKVNPENLVKCFEGQARKOSRHE 180  
 DB 113 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTSSVKVNPENLVKCFEGQARKOSRHE 172  
 QY 181 YOLGFTLIMKNVPTQMKFSVQTMCPLEGSGNIAEFLFSIFGQKHNAAVTLTLDINSV 240  
 DB 173 YOLGFTLIMKNVPTQMKFSVQTMCPLEGSGNIAEFLFSIFGQKHNAAVTLTLDINSV 232  
 QY 241 MFOLREGSSKEKAAVFRSNMISALGRSPMLVGNELTVADVVLMSVLQOTGSSGAAPTNO 300  
 DB 233 IFOLREGSSKEKAAVFRSNMISALGRSPMLVGNELTVADVVLMSVLQOTGSSGAAPTNO 291  
 QY 301 RMLKSCENLAPFSTALQLEK 320  
 DB 292 KMKKSCENLAPFSTALQLEK 311

## RESULT 10

ID Q7T3C0 PRELIMINARY; PRT; 321 AA.  
 AC Q7T3C0;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE zgc:63976.  
 GN ORFNames=zgc:63976;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NC NCB1\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

Query	Subject	Score	Length	DB 2	DB 1	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 37
-------	---------	-------	--------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	-------

QY	3	MYGVKRYHGGASAPRLKVELPTCMYRLPNV-----HSKTTSPATDAGHVOETSEBSLOA	54
Db	1	MYELKLT-----LPQFDIKLPTCMYPLKVNLSLAADSIAGSSGTSASTSA---STSSCDDTAS	54
QY	55	LESRODDILKRLYEKKAANDGSKMHIHPDADLVNTINILQADEPTTLATNTLDSVGLK	111
Db	55	VAARQGVKVLKQJEBELKAQDGGQR-----AGLQVCG--KTFOHTTAFOHG-----GL	98
QY	115	DYGLAKDVIYANNPASPELSLVLRLLCERYRVLSTVHTHSSVKNV-----PENLVK	167
Db	99	KEVPLQDDVAVINHPFIYALALAKNMRNLTYTIDVTFTHSTPADIGPAAREFEANLAK	150
QY	168	CFCEQARAKOSRHEVGLFTLLMKNVPTKOMKFSVQTCPIEGRGNIARPLFSLFGQKHNA	222
Db	159	VPNPALP-----KISTVLLMKNCERTENMISPTMYVPLFYGEVNIIRKIRGVGPAEYVY	211
QY	228	VTLT-----IDSKVDIAMEFQLEGGSSKEKAAVFSGMSNALSGRSPWLGNELTVADVLMSV	287
Db	213	EGSPPLCNEIDLVDLCYQLLRCNTHKTQVAMVRLDKRLQKQYFGGQMSVADYGVYSS	277
QY	285	L 285	
Db	273	L 273	
RESULT	13		
Q8T060		PRELIMINARY; PRT; 313 AA.	
AC	Q8T060		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-MAR-2004 (TREMBLrel. 26, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	LD25772P (fragment).		
GN	ORFNames=CG12304;		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydriidae; Drosophilidae; Drosophila.		
OX	NCBI_TextID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkeley.		
RA	Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,		
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,		
RA	George R., Gonzalez M., Guatin H., Kronmiller B., Li P., Liao G.,		
RA	Miranda A., Mungall C.J., Nunco J., Pacled J., Patagas V., Park S.,		
RA	Patel S., Phouanvong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,		
RA	Celniker S.,		
RL	Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY069537; AL39682.2; -		
DR	FLYbase; FBgn0036515; CG12304.		
DR	InterPro; IPR004046; GST_Cterm.		
DR	InterPro; IPR010987; GST_C-like.		
DR	Pfam; PF00043; GST_C, 1.		
FT	NON TER		
SQ	SEQUENCE 313 AA; 34522 MW; 5AF9C801877F8531 CRC64;		
Query Match	13.5%; Score 224; DB 2; Length 313;		
Best Local Similarity	28.6%; Pred. No. 4,3e-10;		
Matches	86; Conservative 47; Mismatches 122; Indels 46; Gaps 9		
QY	3	MYGVKRYHGGASAPRLKVELPTCMYRLPNV-----HSKTTSPATDAGHVOETSEBSLOA	54
Db	13	MYELKLT-----LPQFDIKLPTCMYPLKVNLSLAADSIAGSSGTSASTSA---STSSCDDTAS	66
QY	55	LESRODDILKRLYEKKAANDGSKMHIHPDADLVNTINILQADEPTTLATNTLDSVGLK	114
Db	67	VAARQGVKVLKQJEBELKAQDGGQR-----AGLQVCG--KTFOHTTAFOHG-----GL	110
QY	115	DYGLAKDVIYANNPASPELSLVLRLLCERYRVLSTVHTHSSVKNV-----PENLVK	167
Db	111	KEVPLQDDVAVINHPFIYALALAKNMRNLTYTIDVTFTHSTPADIGPAAREFEANLAK	170

```
QY 168 CFEQARQKSRHEYOGLFTLLIWKVNPXTOMKFSVQTMCEPIEGSGNIARFLFSLFGQKINA 227
DB 171 VFNVPALP-----KISVTLLWKNCNHTMISSPTMYVPIYGEVNIIRILGVRGPRERY 224
QY 228 VTLTL---IDSWVDIAMFOLREGSSKEKAIVFRSNMNSALGRSPWLVEGNETLTADVVLMGV 284
DB 225 EGSPICNEIDVLIDICYQLLRKNTHTKYAVMRLDLKRLQKQOYRFGSGSMVADYGVYSS 284
QY 285 L 285
DB 285 L 285

RESULT 14
Q6NKM4 PRELIMINARY; PRT; 322 AA.
ID Q6NKM4
AC Q6NKM4;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Lp12114p (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celinker S.;
RU Submitted (May-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BT012670; AAT08476.1; -
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1.
FT NON TER
SQ SEQUENCE 322 AA; 35528 MW; 82A15927DF928DAE CRC64;

Query Match 13.5%; Score 224; DB 2; Length 322;
Beet Local Similarity 28.6%; Pred. No. 4,5e-10;
Matches 86; Conservative 47; Mismatches 122; Indels 46; Gaps 9;

QY 3 MYQVPRYHGSGAPLRYELPTCMYRLPNV-----HSKTSPTATDAGHVQSEPSLQA 54
DB 22 MELTLTL---LPQFDIKLPTCMYRLPNVSLADSLASGSSASTA---STSSCDTAS 75
QY 55 LESRODDILKRLYEKKAANDGSKMHTPDADLDVTNIIQADEPTLATNTLDLNSVLGK 114
DB 76 VAAREBKVLKQLELKAQOQGR-----AGLGVC--KTFQHTTAFQNG-----GL 119
QY 115 DYGLAKDIYINANPASPPLSLILVHRLLCERYRVLSVTHSSVKNV-----PENLVK 167
DB 120 KEVLPQDVVINGHPNFIYPALLALAKNAMRLTYTIDKFTFHTMADIGPARAREFEANLAK 179
QY 168 CFEQARQKSRHEYOGLFTLLIWKVNPXTOMKFSVQTMCEPIEGSGNIARFLFSLFGQKINA 227
DB 180 VFNVPALP-----KISVTLLWKNCNHTMISSPTMYVPIYGEVNIIRILGVRGPRERY 233
QY 228 VTLTL---IDSWVDIAMFOLREGSSKEKAIVFRSNMNSALGRSPWLVEGNETLTADVVLMGV 284
DB 234 EGSPICNEIDVLIDICYQLLRKNTHTKYAVMRLDLKRLQKQOYRFGSGSMVADYGVYSS 293
QY 285 L 285
DB 294 L 294

RESULT 15
MCA2_DROME STANDARD; PRT; 334 AA.
AC Q9VUR3;
DT 16-OCT-2001 (Rel. 40, Created)
```

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Probable multisynthetase complex auxiliary component p38.
CN ORFNames=CG12304;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoeking R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milio G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa K.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kamnitsis J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kratz C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Pacleb J.M.,
RA Palazzolo M., Picman G.S., Pan S., Pollard J., Put V., Reese M.G.,
RA Reihert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splitter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN [2]
RP GENOME REANNOTATION.
RX MEDLINE=24246069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shy S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -!- FUNCTION: Probable core protein of the multisynthetase complex
CC that serves as a template for the assembly of the supramolecular
CC structure (By similarity).
CC -!- SUBUNIT: Component of the multisynthetase complex which is
CC comprised of a bifunctional glutamyl-prolyl-trna synthetase, the
CC monospecific isoleucyl-, leucyl-, glutamyl-, methionyl-, lysyl-,
CC arginyl-, and asparaginyl-trna synthetases as well as three auxiliary
CC proteins, p18, p48 and p43 (By similarity).
CC -!- SIMILARITY: Contains 1 GST-like domain.
```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ch).
CC -----
DR EMBL: AE003530; AAF49612.1; -.
DR IntAct: Q9VUR3; -.
DR Flybase: FBgn0036515; CG12304.
DR InterPro: IPR010987; GST_C-like.
DR InterPro: IPR004046; GST_C-term.
DR Pfam: PF00043; GST_C; 1.
DR KW Protein biosynthesis.
FT DOMAIN 280
SQ SEQUENCE 334 AA; 36933 MW; B68FD70AE62190F CRC64;

Query Match 13.1%; Score 216.5; DB 1; Length 334;
Best Local Similarity 26.0%; Pred. No. 26-09;
Matches 86; Conservative 48; Mismatches 124; Indels 73; Gaps 9;

QY 3 MYQYKPYHGSAPLRVLPFCMRLPNV-----HSKTTSPATDAG----- 42
Db 1 MYELKTL--LPQFDIKLPTCMYPLKNVSLAADSLASGSSTASTSCKLEANNRID 57
QY 43 -----HVQETSEPSLOAESRODILKRYELKAANDGLSKMHTPD 84
Db 58 RTGNMNAATCALDLSLGRQIQRLIKDPTASVAAEQEKVLEKQLELKAQLGQIR----- 110
QY 85 ADDLVNIILOADEPTLATNTLDINSVYLGKDYALNDIYINANPASPLSLVLRHLCE 144
Db 111 AGGVCG--KTFQHTTAFQNG-----GLKEVPLQDVVINGHNPFIYALLALKNAMRN 161
QY 145 RYRYLSTVHTHSSVKNV-----PENVYKCFGEQARKOSRHEVYQUGFTLIMKNVFKTOM 197
Db 162 LYTIIDVKTFTHSTWADIGPAAREFEANLAKVPVPALP-----KISVTLIMKNCETEM 215
QY 198 KFSVQTCPIEGGNIAFLFSLFGQKHNAVTLTL--IDSWVDIAMFQLRGSSSEKXA 254
Db 216 ISSPTWVPIYGEVNIIRYLGAVGPAEYREGSPLCNEIDVIDICYQLRCNTHKTQYA 275
QY 255 VFSNMSALGRSPWLVGNELTVADVVLKSVL 285
Db 276 MVRLLDKRLQKQYFGSQMSVADGVYSSL 306

```

Search completed: February 23, 2005, 13:57:14  
 Job time : 62.9811 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:25:39 / Search time 70.371 Seconds  
(without alignments)  
1758.725 Million cell updates/sec

Title: US-10-622-817-6

Perfect score: 1655

Sequence: 1 MPMYGVKRYHGSAPLRYEL.....RWLKSCENLAPSTALQLLK 320

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp19808:\*  
2: geneseqp19908:\*  
3: geneseqp20008:\*  
4: geneseqp20018:\*  
5: geneseqp20028:\*  
6: geneseqp20038:\*  
7: geneseqp20038:\*  
8: geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1464	88.5	320	7	AD588613 Human Pro
2	1464	88.5	320	8	ABM80843 Tumour-ab
3	1436	86.8	312	2	AAW25776 JTV1 prot
4	1436	86.8	312	8	ADR86551 1-312 am
5	1232	74.4	272	7	ADC10204 Human NOV
6	1069	64.6	229	8	ADR86553 84-312 am
7	735	44.4	161	8	ADR86552 1-161 am
8	248	15.0	51	5	ABP01502 Human ORF
9	216.5	13.1	334	4	AB862468 Drosophila
10	148	8.9	263	4	ABG23964 Novel hum
11	126.5	7.6	716	3	AAQ1411 Arabidops
12	126.5	7.6	719	3	AAQ1410 Arabidops
13	126.5	7.6	748	3	AAQ1409 Arabidops
14	106	6.4	1512	4	AAU04349 Mammalian
15	106	6.4	1512	4	AAW78732 Human pro
16	106	6.4	1550	4	AAW79716 Human pro
17	104	6.3	980	6	ADN23683 Bacterial
18	102	6.2	1049	8	ABM67090 Phototrab
19	101.5	6.1	724	4	AAQ85046 Shrimp wh
20	99	6.0	691	6	ABU16445 Proteins e
21	98	5.9	243	4	AAU58158 Propionib
22	98	5.9	243	6	ABM54677 Propionib
23	98	5.9	644	6	ABJ18952 Pathogen
24	98	5.9	690	4	AAU34345 Staphyloc
25	98	5.9	690	6	ABM71749 Staphyloc

26	98	5.9	985	8	ADN88622	Adn88622 Rat epid1
27	98	5.9	997	8	ADN88620	Adn88620 Rat epid1
28	98	5.9	1004	8	ADN88621	Adn88621 Rat epid1
29	98	5.9	1013	8	ADN88624	Adn88624 Rat epid1
30	98	5.9	1016	8	ADN88619	Adn88619 Rat epid1
31	98	5.9	1032	8	ADN88618	Adn88618 Rat epid1
32	96.5	5.8	1398	8	ABM83058	Abm83058 Human dia
33	96.5	5.8	1398	8	ABM83057	Abm83057 Human dia
34	96.5	5.8	1440	8	ADN03674	Adn03674 Herpeticid
35	96.5	5.8	1440	8	ADP54124	Adp54124 Human PRO
36	96.5	5.8	1440	8	ADP23041	Adp23041 PRO polyp
37	96	5.8	2273	2	AAQ88811	AAQ88811 Erythiphe
38	95	5.7	722	8	ADS20980	Ads20980 Bacterial
39	94	5.7	436	5	ABBS4068	Abbs4068 Lactococc
40	93.5	5.6	1049	5	ABR91239	Abbr91239 Herpeticid
41	93.5	5.6	1064	5	AAE25162	AAE25162 RCH1.5 pr
42	93.5	5.6	1090	4	ABG17694	Abg17694 Novel hum
43	93.5	5.6	1090	4	ABG18311	Abg18311 Novel hum
44	93	5.6	204	8	ADS21772	AdS21772 Bacterial
45	93	5.6	559	2	AAK10682	AaK10682 Polyhydro

#### ALIGNMENTS

RESULT 1  
ID AD588613 standard; protein; 320 AA.  
XX  
AC AD588613;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein Q13155, SEQ ID NO 4489.  
XX  
XX Human; pain; neuronal tissue; gene therapy;  
KW Human; spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spinal segmental nerve injury; SNR; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
XX  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO) GEN HOSPITAL CORP.  
PA (FARB) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; Q13155.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 320 AA;

Query Match 88.5%; Score 1464; DB 7; Length 320;  
Best Local Similarity 87.2%; Pred. No. 7, 5e-142;  
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MPWYQVKPYHGGGAPRLVELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSLQALESRD 60  
DB 1 MPWYQVKPYHGGGAPRLVELPTCMYRLPNVHGRSGYPAPGAGHVGESNLSLQALESRD 60  
QY 61 DILKRLYEKAAVDGSKMHTPDADLVNTIIQADEPTTLATNTLDNSVLGKDYGALK 120  
DB 61 DILKRLYEKAAVDGSKMHTPDADLVNTIIQADEPTTLATNTLDNSVLGKDYGALK 120  
QY 121 DIVINANPASPLSLVLRHLCERYVLSTVHTSSVKNVBNENLKCFCGEQKOSRHE 180  
DB 121 DIVINANPASPLSLVLRHLCERYVLSTVHTSSVKNVBNENLKCFCGEQKOSRHE 180  
QY 181 YOLGFTLIWKNVPKTOMKFSVQTMCPIDEGBNIAFLFSLFGKHNAAVTLTIDSWVDIA 240  
DB 181 YOLGFTLIWKNVPKTOMKFSVQTMCPIDEGBNIAFLFSLFGKHNAAVTLTIDSWVDIA 240  
QY 241 MFOLRGSSKEKAAVFRSNMSALGKSPMLVGNELTVADVLMVSVOQTGGSSGAAPTNOQ 300  
DB 241 IFOLRGSSKEKAAVFRSNMSALGKSPMLVGNELTVADVLMVSVOQTGGSSGAAPTNOQ 300  
QY 301 RMLKSCENLAPFSTALQLLK 320  
DB 301 RMRSCENLAPFNTALKLK 320

RESULT 2  
ABM80843  
ID ABM80843 standard; protein; 320 AA.

AC ABM80843;

DT 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) polypeptide PRO81501, SEQ.2179.

KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.

OS Homo sapiens.

XX MO2004030615-A2.

PN 15-APR-2004.

PD

XX 29-SEP-2003; 2003MO-US028547.  
PF  
XX  
XX 02-OCT-2002; 2002US-0414971P.  
PR  
XX  
XX (GENH) GENENTECH INC.  
PA  
XX  
XX Wu TD, Zhang Z, Zhou Y;  
PI  
XX  
XX WPI; 2004-347921/32.  
DR  
XX  
XX N-PSDB; ACN38579.  
PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
XX  
XX Claim 12; SEQ ID NO 2179; 7273bp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acid and polypeptide  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention

XX Sequence 320 AA;

Query Match 88.5%; Score 1464; DB 8; Length 320;  
Best Local Similarity 87.2%; Pred. No. 7, 5e-142;  
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MPWYQVKPYHGGGAPRLVELPTCMYRLPNVHSKTTSPATDAGHVOETSEPSLQALESRD 60  
DB 1 MPWYQVKPYHGGGAPRLVELPTCMYRLPNVHGRSGYPAPGAGHVGESNLSLQALESRD 60  
QY 61 DILKRLYEKAAVDGSKMHTPDADLVNTIIQADEPTTLATNTLDNSVLGKDYGALK 120  
DB 61 DILKRLYEKAAVDGSKMHTPDADLVNTIIQADEPTTLATNTLDNSVLGKDYGALK 120  
QY 121 DIVINANPASPLSLVLRHLCERYVLSTVHTSSVKNVBNENLKCFCGEQKOSRHE 180  
DB 121 DIVINANPASPLSLVLRHLCERYVLSTVHTSSVKNVBNENLKCFCGEQKOSRHE 180  
QY 181 YOLGFTLIWKNVPKTOMKFSVQTMCPIDEGBNIAFLFSLFGKHNAAVTLTIDSWVDIA 240  
DB 181 YOLGFTLIWKNVPKTOMKFSVQTMCPIDEGBNIAFLFSLFGKHNAAVTLTIDSWVDIA 240  
QY 241 MFOLRGSSKEKAAVFRSNMSALGKSPMLVGNELTVADVLMVSVOQTGGSSGAAPTNOQ 300  
DB 241 IFOLRGSSKEKAAVFRSNMSALGKSPMLVGNELTVADVLMVSVOQTGGSSGAAPTNOQ 300  
QY 301 RMLKSCENLAPFSTALQLLK 320  
DB 301 RMRSCENLAPFNTALKLK 320

RESULT 3  
AAW25776



ID AAW25776 standard; protein; 312 AA.  
 XX  
 AC AAW25776;  
 XX  
 DT 19-DEC-1997 (first entry)  
 XX  
 DE JTV1 protein.  
 XX  
 KW JTV1; hPMS2; probe; detection; chromosome 7; deletion;  
 KW mismatch repair gene; hereditary non-polyposis colorectal cancer;  
 KW homologous recombination.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9708312-A1.  
 XX  
 PD 06-MAR-1997.  
 XX  
 PF 26-AUG-1996; 96WO-US013598.  
 XX  
 PR 24-AUG-1995; 95US-00518862.  
 XX  
 PA (UYJO) UNIV JOHNS HOPKINS.  
 XX  
 PI Vogelestein B, Kinzler KW, Nicolaides NC;  
 XX  
 DR WPI; 1997-179269/16.  
 DR N-PSDB; AAT86182.  
 XX  
 PT Novel chromosome 7 gene, JTV1 - used for detecting chromosome 7  
 PT deletions, and PMS2 promoter activity.  
 XX  
 PS Claim 5; Fig 2; 31pp; English.  
 XX  
 CC This sequence is JTV1 protein and is encoded by DNA isolated from human  
 CC chromosome 7. The JTV1 coding sequence is located upstream from hPMS2.  
 CC JTV1 cDNA can be used as probes to detect chromosome 7 deletions  
 CC involving JTV1. Due to the overlapping promoter regions, deletions of  
 CC JTV1 would also affect PMS2 (a mismatch repair gene) expression, leading  
 CC to hereditary non-polyposis colorectal cancer. JTV1 can also be used to  
 CC assay activity or competence of the PMS2 promoter region, the presence of  
 CC JTV1 suggesting that the PMS2 promoter is intact. JTV1 sequences can also  
 CC be used to guide homologous recombination at the PMS2 locus  
 CC  
 XX  
 SQ Sequence 312 AA;  
 Query Match 86.8%; Score 1436; DB 2; Length 312;  
 Best Local Similarity 87.5%; Pred. No. 5,6e-139;  
 Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MPMYVKKRYHGGASALRVELPTCMYRLPNVHNSKTTSPATDAGHVQETSEPSIQALESRD 60  
 DB 1 MPMYVKKRYHGGASALRVELPTCMYRLPNVHNSKTTSPATDAGHVQETSEPSIQALESRD 60  
 QY 61 DILKRLYEKAAVDSLSKMIHTPPDADLDVTNIIQADEPTLTATNTLDLSVVGKDYGALK 120  
 DB 61 DILKRLYEKAAVDSLSKMIHTPPDADLDVTNIIQADEPTLTATNTLDLSVVGKDYGALK 120  
 QY 121 DIVINANPASPPLSLVLRLLCERYRVLSVTHSSVKNVBNLVKCFGEQARKQSRHE 180  
 DB 121 DIVINANPASPPLSLVLRLLCERYRVLSVTHSSVKNVBNLVKCFGEQARKQSRHE 180  
 QY 181 YOLGTTILWKNVPTQMKFSVOTMCPIBEENIARFLPSLFGQKHNANVATLIDSWVDIA 240  
 DB 181 YOLGTTILWKNVPTQMKFSVOTMCPIBEENIARFLPSLFGQKHNANVATLIDSWVDIA 240  
 QY 241 MFOLEGGSSKEKAAVFRSNNSALGRSPMLVGNELTVADVLSVYQQTGGSSGAAPTNYQ 300  
 DB 241 IFOLKEGGSSKEKAAVFRSNNSALGRSPMLVGNELTVADVLSVYQQTGGSSGAAPTNYQ 300  
 QY 301 RWLKSCENLAPP 312  
 DB 301 RWRMSCENLAPP 312

RESULT 4  
 ID ADR86551 standard; protein; 312 AA.  
 XX  
 AC ADR86551;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE 1-312 amino acid sequence of p38/JTV-1 protein.  
 XX  
 KW p38/JTV-1; cytoskeletal; cancer; leukemia; anticancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1454628-A2.  
 XX  
 PD 08-SEP-2004.  
 XX  
 PF 09-SEP-2003; 2003EP-00020344.  
 XX  
 PR 03-MAR-2003; 2003KR-00013058.  
 XX  
 PA (UYSE-) UNIV SEOUL NAT IND FOUND.  
 XX  
 PI Kim S, Park B;  
 XX  
 DR WPI; 2004-627822/61.  
 DR N-PSDB; ADR86548.  
 XX  
 PT New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
 PT e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
 PT and cutaneous or intraocular melanoma, as well as for screening new  
 PT anticancer agents.  
 XX  
 PS Claim 5; SEQ ID NO 4; 47pp; English.  
 XX  
 CC The present invention relates to an isolated p38/JTV-1 protein for use as  
 CC medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
 CC useful as medicament for treating breast cancer, large intestinal cancer,  
 CC lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
 CC cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
 CC cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
 CC rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
 CC endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
 CC Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
 CC cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
 CC tumour, urethral cancer, penile cancer, prostate cancer, chronic or acute  
 CC leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
 CC cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
 CC CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary  
 CC adenoma, or their combination. The protein is useful as a target for  
 CC screening new anticancer agents. The present sequence represents the 1-  
 CC 312 amino acid sequence of p38/JTV-1 protein.  
 CC  
 XX  
 SQ Sequence 312 AA;  
 Query Match 86.8%; Score 1436; DB 8; Length 312;  
 Best Local Similarity 87.5%; Pred. No. 5,6e-139;  
 Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MPMYVKKRYHGGASALRVELPTCMYRLPNVHNSKTTSPATDAGHVQETSEPSIQALESRD 60  
 DB 1 MPMYVKKRYHGGASALRVELPTCMYRLPNVHNSKTTSPATDAGHVQETSEPSIQALESRD 60  
 QY 61 DILKRLYEKAAVDSLSKMIHTPPDADLDVTNIIQADEPTLTATNTLDLSVVGKDYGALK 120  
 DB 61 DILKRLYEKAAVDSLSKMIHTPPDADLDVTNIIQADEPTLTATNTLDLSVVGKDYGALK 120  
 QY 121 DIVINANPASPPLSLVLRLLCERYRVLSVTHSSVKNVBNLVKCFGEQARKQSRHE 180  
 DB 121 DIVINANPASPPLSLVLRLLCERYRVLSVTHSSVKNVBNLVKCFGEQARKQSRHE 180

```
QY 181 YOLGFTLWKVPTOMKFSVOTMCPFBEGNIAFLFSLFGQHNATLTLINSWDIA 240
DB 181 YOLGFTLWKVPTOMKFSVOTMCPFBEGNIAFLFSLFGQHNATLTLINSWDIA 240
QY 241 MFOLEGGSSKEKAAVFRSMNSALGRSPWLVGNETLVADVLMSTLQOTGGSSGAAPTNVQ 300
DB 241 IFOLKEGSSKEKAAVFRSMNSALGRSPWLVGNETLVADVLMSTLQOTGGSSGAAPTNVQ 300
QY 301 RMLKSCENLAPF 312
DB 301 RMLKSCENLAPF 312

RESULT 5
ADCL0204
ID ADCL0204 standard; protein; 272 AA.
AC ADCL0204;
DT 18-DEC-2003 (first entry)
XX Human NOVX polypeptide SEQ ID NO: 226.
DE
XX
KM cytostatic; antidiabetic; anorectic; cerebroprotective; neuroprotective;
KW antiinflammatory; gene therapy; antisense therapy; thymimetic; NOVX;
KW pathology; cancer; diabetes; obesity; endocrine disorder; CNS disorder;
KW inflammatory disorder; chromosome mapping; tissue typing;
XX predictive medicine.
XX
OS Homo sapiens.
XX
PN WO2003000842-A2.
XX
PD 03-JAN-2003.
XX
PF 04-JUN-2002; 2002WO-US017443.
XX
PR 04-JUN-2001; 2001US-0295607P.
PR 04-JUN-2001; 2001US-0295610P.
PR 06-JUN-2001; 2001US-0296404P.
PR 06-JUN-2001; 2001US-0296418P.
PR 07-JUN-2001; 2001US-0296575P.
PR 11-JUN-2001; 2001US-0297414P.
PR 12-JUN-2001; 2001US-0295573P.
PR 12-JUN-2001; 2001US-0297567P.
PR 14-JUN-2001; 2001US-0296285P.
PR 15-JUN-2001; 2001US-0296285P.
PR 18-JUN-2001; 2001US-0299133P.
PR 19-JUN-2001; 2001US-0299230P.
PR 21-JUN-2001; 2001US-0299949P.
PR 22-JUN-2001; 2001US-0300177P.
PR 26-JUN-2001; 2001US-0300883P.
PR 28-JUN-2001; 2001US-0301530P.
PR 28-JUN-2001; 2001US-0301530P.
PR 03-JUL-2001; 2001US-0302951P.
PR 31-JUL-2001; 2001US-0308890P.
PR 14-SEP-2001; 2001US-0322297P.
PR 25-SEP-2001; 2001US-0324659P.
PR 03-DEC-2001; 2001US-0337477P.
PR 14-DEC-2001; 2001US-0341562P.
PR 21-FEB-2002; 2002US-0358656P.
PR 21-FEB-2002; 2002US-0359122P.
PR 22-FEB-2002; 2002US-0358978P.
PR 22-FEB-2002; 2002US-0359034P.
PR 22-FEB-2002; 2002US-0359035P.
PR 27-FEB-2002; 2002US-0359121P.
PR 01-MAR-2002; 2002US-0359964P.
PR 12-MAR-2002; 2002US-0363430P.
PR 12-MAR-2002; 2002US-0363676P.
PR 10-APR-2002; 2002US-0371346P.
PR 10-MAY-2002; 2002US-0379444P.
```

```
PR 04-JUN-2002; 2002US-00379444.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Agee ML, Anderson DM, Bergs C, Casman SJ, Catterton E;
PI DiPippo VA, Edinger SR, Eissen A, Ellerman K, Gangoli EA;
PI Gerlach VL, Gorman L, Guo X, Hermann JM, Hjalte T, Ji W, Kekuda R;
PI Khramtsov NV, Li L, Liu X, Malyanar UM, Miller CE, Miller I;
PI Ort T, Padigaru M, Patturajan M, Pena CE, Rastelli L, Rieger DK;
PI Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G, Spaderna SK;
PI Spytek KA, Stone DJ, Vernet CAM, Zhong H, Zhong W, Alsbrook JP;
PI Burgess CE, Lepley DM;
XX
XX WPI; 2003-210149/20.
XX
XX N-PSDB; ADCL0203.
XX
PT New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
XX
PS Claim 1; SEQ ID NO 226; 772pp; English.
XX
XX
CC The invention relates to novel isolated polypeptides, mature form of the
CC polypeptide, a sequence that is 95% identical to the polypeptide or the
CC polypeptide comprising one or more conservative substitutions. The NOVX
CC polypeptide is useful for treating or preventing a pathology associated
CC with the polypeptide e.g. disorders associated with aberrant expression
CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC endocrine, CNS and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. This sequence corresponds to one of the
CC polypeptides of the invention.
XX
XX
SQ Sequence 272 AA;
XX
XX
Query Match 74.4%; Score 1232; DB 7; Length 272;
Best Local Similarity 88.4%; Pred. No.5.1e-116;
Matches 237; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
XX
QY 45 QETSEPSLOAESRQDDILKRLYEKAADVGLSKMHTPDADLDVNTLQADEPTLTATN 104
DB 5 QBSNSLSTQALBSRQDDILKRLYEKAADVGLSKMHTPDADLDVNTLQADEPTLTATN 64
XX
QY 105 TLDLNSVLGKDYALDVIYNNPASPPLSLVYHRLCERFRVYSTVHTSSVKNVPE 164
DB 65 ALDPLNSVLGKDYALDVIYNNPASPPLSLVYHRLCERFRVYSTVHTSSVKNVPE 124
XX
QY 165 LVKCFGEQARKOSRHEVQLGFTLWKVPTOMKFSVOTMCPFBEGNIAFLFSLFGQK 224
DB 125 LTKCFGEQARKOPRODYQLGFTLWKVPTOMKFSVOTMCPFBEGNIAFLFSLFGQK 184
XX
QY 225 HNAVNTLTLDSWVDIAMFOLREKSSKEKAAVFRSMNSALGRSPWLVGNETLVADVLM 284
DB 185 HNAVNTLTLDSWVDIAIFOLREKSSKEKAAVFRSMNSALGRSPWLVGNETLVADVLM 244
XX
QY 285 LQOTGGSSGAAPTNVORMLKSCENLAPF 312
DB 245 LQOTGGSSGAAPTNVORMLKSCENLAPF 272
XX
RESULT 6
ADCL0204
ID ADCL0204 standard; protein; 229 AA.
AC ADCL0204;
DT 18-NOV-2004 (first entry)
XX
XX 84-312 amino acid sequence of p38/JTV-1 protein.
DE
XX p38/JTV-1; cytostatic; cancer; leukemia; anticancer.
XX
```

OS Homo sapiens.  
XX  
XX EPI454628-A2.  
XX  
XX 08-SEP-2004.  
XX  
XX 09-SEP-2003; 2003EP-00020344.  
XX  
XX 03-MAR-2003; 2003KR-00013058.  
XX  
XX (UYSE-) UNIV SEOUL NAT IND FOUND.  
XX  
XX Kim S, Park B;  
XX  
XX WPI; 2004-627822/61.  
XX N-PSDB; ADR86550.  
XX  
XX New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
XX e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
XX PT and cutaneous or intraocular melanoma, as well as for screening new  
XX PT anticancer agents.

PS Claim 5; SEQ ID NO 6; 47bp; English.

CC The present invention relates to an isolated p38/JTV-1 protein for use as  
CC medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
CC useful as medicament for treating breast cancer, large intestinal cancer,  
CC lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
CC cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
CC cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
CC rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
CC endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
CC Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
CC cancer, thyroid cancer, parathyroid cancer, adrenal cancer, soft tissue  
CC tumour, urethral cancer, penile cancer, prostate cancer, chronic or acute  
CC leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
CC cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
CC CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary  
CC adenoma, or their combination. The protein is useful as a target for  
CC screening new anticancer agents. The present sequence represents the 84-  
CC 312 amino acid sequence of p38/JTV-1 protein.

XX Sequence 229 AA;

Query Match 64.6%; Score 1069; DB 8; Length 229;

Best Local Similarity 88.2%; Pred. No. 2,6e-101;

Matches 202; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

QY 84 DADLDVTNIIQADEPTTLTNTLDNSVVGKDYGALKDVIINANPASPPLSLTVLHRLLC 143  
DB 1 DADLDVTNIIQADEPTTLTNTLDNSVVGKDYGALKDVIINANPASPPLSLTVLHRLLC 60  
QY 144 ERYRLSTVHTHSVKNVPENLVKCFGEQARKOSRHEVQLGFTLIWKVNPKTOMKFSVOT 203  
DB 61 EHFRLSTVHTHSVKNVPENLVKCFGEQARKOSRHEVQLGFTLIWKVNPKTOMKFSVOT 120  
QY 204 MCPRIEGEGNIAFLSLRFGQKNAATLTLLIDSWNDIAMQLRESSSKKGAAYFRSMNSL 263  
DB 121 MCPRIEGEGNIAFLSLRFGQKNAATLTLLIDSWNDIAMQLRESSSKKGAAYFRSMNSL 180  
QY 264 GRSPLVAGNELTVADVLMVLSVLOQTGSSGGAFTVVRKLSKSCENLAFF 312  
DB 101 GRSPLVAGNELTVADVLMVLSVLOQTGSSGGAFTVVRKLSKSCENLAFF 229

RESULT 7  
ADR86552  
ID ADR86552 standard; protein; 161 AA.  
XX  
XX ADR86552;  
XX  
XX 18-NOV-2004 (first entry)  
XX

DE 1-161 amino acid sequence of p38/JTV-1 protein.

XX p38/JTV-1; Cytostatic; cancer; leukemia; anticancer.

XX Homo sapiens.

XX EPI454628-A2.

XX 08-SEP-2004.

XX 09-SEP-2003; 2003EP-00020344.

XX 03-MAR-2003; 2003KR-00013058.

XX (UYSE-) UNIV SEOUL NAT IND FOUND.

XX Kim S, Park B;

XX WPI; 2004-627822/61.

XX N-PSDB; ADR86549.

PT New isolated p38/JTV-1 protein, useful as medicament for treating cancer  
PT e.g., stomach, liver, blood, bone, pancreatic, skin, head or neck cancer  
PT PT and cutaneous or intraocular melanoma, as well as for screening new  
PT anticancer agents.

PS Claim 5; SEQ ID NO 5; 47bp; English.

CC The present invention relates to an isolated p38/JTV-1 protein for use as  
CC medicament. The p38/JTV-1 protein or the pharmaceutical composition is  
CC useful as medicament for treating breast cancer, large intestinal cancer,  
CC lung cancer, small cell lung cancer, stomach cancer, liver cancer, blood  
CC cancer, bone cancer, pancreatic cancer, skin cancer, head or neck cancer,  
CC cutaneous or intraocular melanoma, uterine sarcoma, ovarian cancer,  
CC rectal cancer, anal cancer, colon cancer, fallopian tube carcinoma,  
CC endometrial carcinoma, cervical cancer, vulval cancer, vaginal carcinoma,  
CC Hodgkin's disease, esophageal cancer, small intestine cancer, endocrine  
CC cancer, thyroid cancer, parathyroid cancer, adrenal cancer, chronic or acute  
CC leukemia, lymphocytic lymphoma, bladder cancer, kidney cancer, ureter  
CC cancer, renal cell carcinoma, renal pelvic carcinoma, CNS tumour, primary  
CC CNS lymphoma, bone marrow tumour, brain stem nerve gliomas, pituitary  
CC adenoma, or their combination. The protein is useful as a target for  
CC screening new anticancer agents. The present sequence represents the 1-  
CC 161 amino acid sequence of p38/JTV-1 protein.

XX Sequence 161 AA;

Query Match 44.4%; Score 735; DB 8; Length 161;

Best Local Similarity 88.8%; Pred. No. 4.4e-67;

Matches 143; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 MEMYGVKPYHGGASAPLRYELPTCMYRLPNVHSKTTSPATDAGHVBETSEPSIQALESROD 60  
DB 1 MEMYGVKPYHGGASAPLRYELPTCMYRLPNVHSKTTSPATDAGHVBETSEPSIQALESROD 60  
QY 61 DILKRLYELKAAVNDLSKVIITPPDLDVTNIIQADEPTTLTNTLDNSVVGKDYGALK 120  
DB 61 DILKRLYELKAAVNDLSKVIITPPDLDVTNIIQADEPTTLTNTLDNSVVGKDYGALK 120  
QY 121 DIVINANPASPPLSLTVLHRLLCERVRVLTSTHTSSVKNV 161  
DB 121 DIVINANPASPPLSLTVLHRLLCERVRVLTSTHTSSVKNV 161

RESULT 8  
ABP01502  
ID ABP01502 standard; protein; 51 AA.  
XX  
XX ABP01502;  
XX  
XX 24-JUN-2002 (first entry)  
XX

DE Human ORFX protein sequence SEQ ID NO:2986.  
XX Human, open reading frame; ORFX, gene therapy; cancer; cirrhosis;  
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;  
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
XX hyperextension; hypothyroidism; cholesterol ester storage disease;  
XX immune deficiency; immune disorder; infectious disease;  
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
XX myasthenia gravis.  
XX Homo sapiens.  
OS WO200192523-A2.  
XX  
XX  
XX PD 06-DEC-2001.  
XX  
XX PF 29-MAY-2001; 2001WO-US010836.  
XX  
XX PR 30-MAY-2000; 2000US-0206132P.  
XX PR 29-AUG-2000; 2000US-0228716P.  
XX  
XX (CUBA-) CUBAGEN CORP.  
XX  
XX Shinkets RA, Leach MD;  
PI WPI; 2002-106308/14.  
XX DR N-PSDB; ABN17254.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
XX Disclosure; SEQ ID NO 2986; 1037P; English.  
XX  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders related to organ  
CC osteoarthritis, neurodegenerative diseases, diabetes mellitus, systemic  
CC transplantion, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune inflammatory eye disease, graft-versus-host  
CC disease and autoimmune inflammatory diseases. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 51 AA;  
Query Match 15.0%; Score 248; DB 5; Length 51;  
Best Local Similarity 96.1%; Pred. No. 1.3e-11;  
Matches 49; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 9  
ABB62468

ID ABB62468 standard; protein; 334 AA.  
XX  
XX ABB62468;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster polypeptide SEQ ID NO 14196.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li FWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX DR N-PSDB; ABL06571.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 14196; 21P + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 334 AA;  
Query Match 13.1%; Score 216.5; DB 4; Length 334;  
Best Local Similarity 26.0%; Pred. No. 4.8e-13;  
Matches 86; Conservative 48; Mismatches 124; Indels 73; Gaps 9;  
QY 3 MYQVRYHGSAPLREVELPTCMYRLPNV-----HSKTTSPATDAG----- 42  
DB 1 MYELTTL---LPQFDIKLPTCMYRLPNVSLADSLASGSSSTASSTSSCKLEANRID 57  
QY 43 -----HVGTSRPSLOALRESRODILKRLYELEKAAVDGLSKMHTPD 84  
DB 58 RTGRNAATCAALDLSLGRIOQLKODTASVAARBEKVLKQLEIKALDGLGIR----- 110  
QY 85 ADLDVTNIIQADEPTTLATNTIDLNSVLGKDYGALGDIIVANPASPPLSLIVLRLLCE 144  
DB 111 AGLGVCQ--KTFQHTTAPQNG-----GLKEVPADVDVINGHFPIYALALAKNAWRN 161  
QY 145 RYRVLSIVTHSSVKNV-----PENLVKCGEQARKQSREHYOLGFTLIMKNVPTQM 197  
DB 162 LYTIDVKTFTHTSTMDIGPARREFEKNLAKYVNPALP-----KISVTLIKKECHHEM 215  
QY 198 KPSVQVMCIIEBEGNARFLFSLFGQKNAVLTLT---IDSWVDIAMFQLRBSSSEKXA 254  
DB 216 ISSPTMVPPIYGVNIRYLGAVGPAEYRIBSSPLCNELDVLDCYQLRKNTHKTQVA 275  
QY 255 VFRSNNSALGRSPWLVGNELTVADVVLMSVL 285

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: February 23, 2005, 13:36:20 ; Search time 17.81 Seconds  
(without alignments)  
1341.256 Million cell updates/sec

Title: US-10-622-817-6

Perfect score: 1655

Sequence: 1 MPWQVKKHYHGSAFLRVEL.....RWLKSCEMLAPSTALQLLK 320

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTCUTS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1464	88.5	341	4 US-09-949-016-11312	Sequence 11312, A
2	1436	86.8	312	2 US-08-518-862C-2	Sequence 2, Appl
3	106	6.4	1512	3 US-09-443-184-48	Sequence 48, Appl
4	96.5	5.8	1440	3 US-09-357-251-37	Sequence 37, Appl
5	90.5	5.5	359	3 US-09-540-824-2	Sequence 2, Appl
6	90	5.4	559	2 US-08-756-317-10	Sequence 10, Appl
7	90	5.4	559	4 US-09-091-609-4	Sequence 4, Appl
8	90	5.4	2954	4 US-09-150-867-1	Sequence 1, Appl
9	90	5.4	5215	3 US-09-105-537-2	Sequence 2, Appl
10	89	5.4	1056	4 US-09-595-684B-29	Sequence 29, Appl
11	89	5.4	1057	3 US-09-541-782-10	Sequence 10, Appl
12	89	5.4	1057	4 US-09-723-820-10	Sequence 10, Appl
13	89	5.4	1287	4 US-10-270-085-10	Sequence 10, Appl
14	89	5.4	1287	4 US-09-949-016-7826	Sequence 7826, Ap
15	88	5.3	712	4 US-09-489-039A-10736	Sequence 10736, A
16	88	5.3	778	4 US-09-583-110-3930	Sequence 3930, Ap
17	88	5.3	786	4 US-09-107-433-3893	Sequence 3893, Ap
18	86.5	5.2	520	4 US-09-248-796A-17644	Sequence 17644, A
19	86.5	5.2	967	4 US-09-540-236-2449	Sequence 2449, Ap
20	86	5.2	535	4 US-09-489-039A-11461	Sequence 11461, A
21	85.5	5.2	556	4 US-09-134-000C-6329	Sequence 6329, Ap
22	85.5	5.2	608	4 US-09-284-768A-4	Sequence 4, Appl
23	85.5	5.2	657	4 US-09-284-768A-7	Sequence 7, Appl
24	85	5.1	222	4 US-09-248-796A-18146	Sequence 18146, A
25	85	5.1	302	4 US-09-107-532A-6924	Sequence 6924, Ap
26	85	5.1	559	4 US-09-672-749-2	Sequence 2, Appl
27	85	5.1	1971	4 US-09-914-272A-1	Sequence 1, Appl

28	85	5.1	1971	4 US-10-638-333-1	Sequence 1, Appl
29	84	5.1	443	4 US-09-949-016-10598	Sequence 10598, A
30	84	5.1	646	4 US-09-248-796A-18678	Sequence 18678, A
31	84	5.1	1057	4 US-09-428-156B-2	Sequence 2, Appl
32	83.5	5.0	360	4 US-09-949-016-10589	Sequence 10589, A
33	83.5	5.0	475	4 US-09-328-352-5379	Sequence 5379, Ap
34	83.5	5.0	519	4 US-09-489-039A-8208	Sequence 8208, Ap
35	83.5	5.0	515	4 US-09-543-681A-6505	Sequence 6505, Ap
36	83.5	5.0	789	3 US-09-002-285-84	Sequence 84, Appl
37	83.5	5.0	789	4 US-09-589-477-84	Sequence 84, Appl
38	83.5	5.0	789	4 US-10-099-285A-84	Sequence 84, Appl
39	83.5	5.0	959	4 US-09-543-681A-6879	Sequence 6879, Ap
40	83	5.0	434	3 US-09-012-072-4	Sequence 4, Appl
41	83	5.0	434	3 US-09-120-601-4	Sequence 4, Appl
42	83	5.0	448	3 US-09-120-601-6	Sequence 6, Appl
43	83	5.0	559	4 US-09-821-016-1	Sequence 1, Appl
44	83	5.0	559	4 US-10-266-787-1	Sequence 1, Appl
45	83	5.0	767	3 US-08-836-567-8	Sequence 8, Appl

## ALIGNMENTS

```
RESULT 1
US-09-949-016-11312
; Sequence 11312, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001037
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11312
; LENGTH: 341
; TYPE: prt
; ORGANISM: Human
US-09-949-016-11312

Query Match      88.5% Score 1464; DB 4; Length 341;
Best Local Similarity 87.2%; Pred. No. 96-153;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY      1 MPWQVKKHYHGSAFLRVELPTQWYRLPNVHSKTTSPATDAGHVOETSPSLQALESRD 60
        |||||
DB      22 MPWQVKKHYHGSAFLRVELPTQWYRLPNVHSKTTSPATDAGHVOETSPSLQALESRD 81
        |||||

QY      61 DLKLKLYEKAAYVDGSKMITHPPDADLVNTNLQADEPTTLATNTLDLSVYGYKGYGAK 120
        |||||
DB      82 DLKLKLYEKAAYVDGSKMITHPPDADLVNTNLQADEPTTLATNTLDLSVYGYKGYGAK 141
        |||||

QY      121 DIVIANPASPPLSLVLRHLCEERYLVSTYHTSSVGVNVPENLVKCGEQAOKOSRAH 180
        |||||
DB      142 DIVIANPASPPLSLVLRHLCEERYLVSTYHTSSVGVNVPENLVKCGEQAOKOSRAH 201
        |||||

QY      181 YOLGFTLWKVNPKTQMKFSVQTMCPBEGEINIAFLFSLFQKKNNAVTLTLIDSVDIA 240
        |||||
DB      202 YOLGFTLWKVNPKTQMKFSVQTMCPBEGEINIAFLFSLFQKKNNAVTLTLIDSVDIA 261
        |||||

QY      241 MQLAEKSKKKAAYFRSNNALGRSPMLVGNELTVADVVLMSVYQQTGSSGGAAPTNY 300
        |||||
DB      262 IFQLKEGSKKKAAYFRSNNALGRSPMLVGNELTVADVVLMSVYQQTGSSGGAAPTNY 321
        |||||

QY      301 RWLKSCEMLAPSTALQLLK 320
```

Db 322 RWRSCENLAPFNTALKLK 341

## RESULT 2

US-08-518-862C-2  
Sequence 2, Application US/08518862C  
Patent No. 5843757  
GENERAL INFORMATION:  
APPLICANT: Vogelstein, Bert  
APPLICANT: Kinzler, Kenneth W.  
APPLICANT: Nicolaides, Nicholas C.  
TITLE OF INVENTION: Human JTV1 Gene Overlaps PMS2 Gene  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: U.S.A.  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,862C  
FILING DATE: 24-AUG-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,49697  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-518-862C-2

Query Match 86.8%; Score 1436; DB 2; Length 312;  
Best Local Similarity 87.5%; Pred. No. 9,6e-150;  
Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

Db 1 MEMYQVKPYHGSAPRLVRLPTCMRLPNVHASKTSPATDAGHVQETSEPSIQALESRQD 60  
1 MEMYQVKPYHGGAPRLVRLPTCMRLPNVHGRSYGAPAGHVOESNTSIQALESRQD 60  
61 DILKRLYELKAANDGISKMIHTPDADLDVTNIIQDEPTLTATNTLIDNSVIGKXYGLK 120  
61 DILKRLYELKAANDGISKMIOTPDADLDVTNIIQDEPTLTATNTLIDNSVIGKXYGLK 120  
121 DIVINANPASPLSLIVHRLICERYVLSTVHTSSVKNVENVKCGEQARKQSRHE 180  
121 DIVINANPASPLSLIVHRLICERYVLSTVHTSSVKNVENVKCGEQARKQSRHE 180  
121 DIVINANPASPLSLIVHRLICERYVLSTVHTSSVKNVENVKCGEQARKQSRHE 180  
181 YQLGFTLWKNVPTOMKFSVOTMCPRIEGGNIAFLSIPGQKKNATLTLIDSWVIA 240  
181 YQLGFTLWKNVPTOMKFSVOTMCPRIEGGNIAFLSIPGQKKNATLTLIDSWVIA 240  
241 IFOLREGSSKEKAIVFRSNGSALGRSPMLVGNELTVADVLMVYLQGTGSSGAAPTVO 300  
241 IFOLREGSSKEKAIVFRSNGSALGRSPMLVGNELTVADVLMVYLQGTGSSGAAPTVO 300  
301 RMLKSCENLAPF 312  
301 RWRSCENLAPF 312

## RESULT 3

US-09-443-184-48  
Sequence 48, Application US/09443184A  
Patent No. 6372431  
GENERAL INFORMATION:  
APPLICANT: Cunningham, Mary Jane  
APPLICANT: Zweiger, Gary  
APPLICANT: Kaser, Matthew R.  
APPLICANT: Panzer, Scott  
APPLICANT: Selhammer, Jeffrey J.  
APPLICANT: Yue, Henry  
APPLICANT: Baughn, Mariah  
APPLICANT: Azimzai, Yalda  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS  
FILE REFERENCE: PC-0007 US  
CURRENT APPLICATION NUMBER: US/09/443,184A  
CURRENT FILING DATE: 1999-11-19  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PERL Program  
SEQ ID NO 48  
LENGTH: 1512  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Inocyte ID No. 6372431 2302721CD1  
US-09-443-184-48

Query Match 6.4%; Score 106; DB 3; Length 1512;

Best Local Similarity 21.8%; Pred. No. 0.096;  
Matches 44; Conservative 33; Mismatches 69; Indels 56; Gaps 8;

Db 124 INANPASPLSLIVHRLICERYVLSTVHTSSVKNVENVKCGEQARKQSRHEYL 183  
6 LTVNSGDPPLGAL-----AVEHVKD-VSISVEEGKENILH----- 41  
184 GFTLWKNVPTOMKFSVOTMCPRIEGGNIAFL-----FSLFGQKKNATLTLIDSW 237  
42 -----VSENVLTVD-----NSILRYIARVATTAAGLYGS--NIMEHTLIDHWL 82  
238 DIAMFOLREGSSKEKAIVFRSNGSALGRSPMLVGNELTVADVLMVYLQ-----QTGSS 292  
83 EFSATKL--SSCSFSTTINELNHCLSLRYYIVGNLSLADLCVMATLKGNAAWQEQLK 140  
293 GAAPTVO RMLKSCENLAPFST 314  
141 KKAPVHVKRWFGLAQAFQS 162

## RESULT 4

US-09-357-251-37  
Sequence 37, Application US/09357251  
Patent No. 6271441  
GENERAL INFORMATION:  
APPLICANT: Falco, S. Carl  
APPLICANT: Farnodu, Layo O.  
APPLICANT: Orozco, Buddy  
APPLICANT: Schwaber, James S.  
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase  
FILE REFERENCE: BB-1193  
CURRENT APPLICATION NUMBER: US/09/357,251  
CURRENT FILING DATE: 1999-07-20  
EARLIER APPLICATION NUMBER: 60/093,530  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 37  
LENGTH: 1440  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-357-251-37

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:57:26 ; Search time 50.2443 Seconds  
(without alignments)  
2084.158 Million cell updates/sec

Title: US-10-622-817-6

Perfect score: 1655

Sequence: 1 MPMYQVKRPHGGSAPLRYEL.....RWLKSCENLAPETALQLLK 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1436	86.8	312	US-10-463-676-4	Sequence 4, Appl1
2	1069	64.6	229	US-10-463-676-6	Sequence 6, Appl1
3	735	44.4	161	US-10-463-676-5	Sequence 5, Appl1
4	105.5	6.4	925	US-10-437-963-126132	Sequence 126132,
5	104	6.3	980	US-10-369-493-6336	Sequence 6336, Ap
6	103.5	6.3	468	US-10-424-599-260388	Sequence 260388,
7	99.5	6.0	914	US-10-437-963-153870	Sequence 153870,
8	99	6.0	691	US-10-282-132A-44369	Sequence 44369, A
9	98	5.9	690	US-09-815-242-5841	Sequence 5841, Ap
10	96.5	5.8	1440	US-09-864-464-37	Sequence 37, Appl
11	96.5	5.8	1440	US-10-786-720-31	Sequence 31, Appl
12	96	5.8	569	US-10-437-963-123904	Sequence 123904,
13	95	5.7	722	US-10-369-493-10013	Sequence 10013, A

14	93.5	5.6	456	15	US-10-425-114-63166	Sequence 63166, A
15	93	5.6	204	15	US-10-369-493-10805	Sequence 10805, A
16	93	5.6	1398	16	US-10-437-963-180124	Sequence 180124,
17	91	5.5	201	15	US-10-369-493-13717	Sequence 13717, A
18	91	5.5	455	15	US-10-425-114-42633	Sequence 42633, A
19	91	5.5	805	15	US-10-424-599-191796	Sequence 191796,
20	91	5.5	1319	16	US-10-437-963-197783	Sequence 197783,
21	91	5.5	1788	16	US-10-437-963-197780	Sequence 197780,
22	90.5	5.5	419	15	US-10-424-599-260387	Sequence 260387,
23	90.5	5.5	578	15	US-10-418-8618-32	Sequence 32, Appl
24	90.5	5.5	661	15	US-10-369-493-2558	Sequence 2558, Ap
25	90.5	5.5	716	15	US-10-369-493-2175	Sequence 2175, Ap
26	90	5.4	559	9	US-09-364-847-21	Sequence 21, Appl
27	90	5.4	856	9	US-09-364-847-33	Sequence 33, Appl
28	90	5.4	856	9	US-09-364-847-35	Sequence 35, Appl
29	90	5.4	1500	16	US-10-437-963-111265	Sequence 111265,
30	90	5.4	5215	9	US-09-861-289-2	Sequence 2, Appl1
31	90	5.4	5215	9	US-09-860-846-2	Sequence 2, Appl1
32	90	5.4	5215	10	US-09-988-3848-2	Sequence 2, Appl1
33	90	5.4	5215	10	US-09-836-821-2	Sequence 2, Appl1
34	90	5.4	5215	14	US-10-271-889-45	Sequence 45, Appl
35	89.5	5.4	203	15	US-10-369-493-8604	Sequence 8604, Ap
36	89	5.4	407	15	US-10-424-599-267113	Sequence 267113,
37	89	5.4	1056	15	US-10-282-174-472	Sequence 472, App
38	89	5.4	1056	15	US-10-282-174-474	Sequence 476, App
39	89	5.4	1056	15	US-10-282-174-476	Sequence 476, App
40	89	5.4	1056	17	US-10-600-009-472	Sequence 474, App
41	89	5.4	1056	17	US-10-600-009-476	Sequence 476, App
42	89	5.4	1056	17	US-10-600-009-476	Sequence 476, App
43	89	5.4	1057	16	US-10-714-796-4	Sequence 4, Appl
44	88.5	5.3	656	15	US-10-369-493-18559	Sequence 18559, A
45	88.5	5.3	799	16	US-10-437-963-112603	Sequence 112603,

## ALIGNMENTS

RESULT 1  
US-10-463-676-4  
; Sequence 4, Application US/10463676  
; Publication No. US20040175375A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Sunghoon  
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method  
; FILE REFERENCE: 012679-091  
; CURRENT APPLICATION NUMBER: US/10/463,676  
; CURRENT FILING DATE: 2003-06-18  
; PRIOR APPLICATION NUMBER: KR 10-2003-13058  
; PRIOR FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)...(312)  
; OTHER INFORMATION: 1-312 amino acid sequence of p38/JTV-1  
US-10-463-676-4  
Query Match 86.8%; Score 1436; DB 16; Length 312;  
Best Local Similarity 87.5%; Pred. No. 2.6e-137;  
Matches 273; Conservative 14; Mismatches 25; Indels 0; Gaps 0;  
QY 1 MPMYQVKRPHGGSAPLRYELPCTCMRLPVHVKTTSPATDAGHVOETSEPSLOAESRQD 60  
DB 1 MPMYQVKRPHGGSAPLRYELPCTCMRLPVHVKTTSPATDAGHVOETSEPSLOAESRQD 60  
QY 61 DILKLEIKAVDGLSKVHTPDADLVNTILQADEPTLATNTLDLSVLGKDYGALK 120

```

Db      61 DILRLVELKAAVDGSKMQLTPDADLVNIIQADEPTTLTNALDLSVLGKDYALK 120
Qy      121 DIVINANPASPPLSLVLRHLRCERYRLSTVHTSSVKVNPENLVYCFEGQARKSRHE 180
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 DIVINANPASPPLSLVLRHLRCERYRLSTVHTSSVKVNPENLVYCFEGQARKSRHE 180
Qy      181 YOLGFTLLIWKVNPPTQMKFSVQVTCPIEGEGNINARPLFSFGQKHNANVTLLIDSWDIA 240
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 YOLGFTLLIWKVNPPTQMKFSVQVTCPIEGEGNINARPLFSFGQKHNANVTLLIDSWDIA 240
Qy      241 MFOLEGGSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVLTQGTGSSGAAPTNNQ 300
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      241 IFOLKESSSKEKAAPFRSMNSALGRSPMLVGNELTVADVLMVLTQGTGSSGAAPTNNQ 300
Qy      301 RMKSCENLAPF 312
        |||:|||||:
Db      301 RMKSCENLAPF 312

```

RESULT 2

```

US-10-463-676-6
; Sequence 6, Application US/10463676
; Publication No. US20040175375A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Sunghoon
; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method
; FILE REFERENCE: 012679-091
; CURRENT APPLICATION NUMBER: US/10/463,676
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: KR 10-2003-13058
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 229
; TYPE: PRP
; ORGANISM: Homo sapiens
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(229)
; OTHER INFORMATION: 84-312 amino acid sequence p38/JTV-1
US-10-463-676-6

```

```

Query Match      64.6%; Score 1069; DB 16; Length 229;
Best Local Similarity 88.2%; Pred. No. 4e-100;
Matches 202; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

Qy      84 DADLVNIIQADEPTTLTNALDLSVLGKDYALKDIVINANPASPPLSLVLRHLRC 143
Db      1 DADLVNIIQADEPTTLTNALDLSVLGKDYALKDIVINANPASPPLSLVLRHLRC 60
Qy      144 ERYVLSVTHSSVKNVPEMLVKCFEGQARKSRHEVQLGFTLLIWKVNPPTQMKFSVQV 203
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 EHFVLSVTHSSVKNVPEMLVKCFEGQARKSRHEVQLGFTLLIWKVNPPTQMKFSVQV 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      204 MCPLEGEGNINARPLFSFGQKHNANVTLLIDSWDIAMFOLEGGSSKEKAAPFRSMNSAL 263
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 MCPLEGEGNINARPLFSFGQKHNANVTLLIDSWDIAMFOLEGGSSKEKAAPFRSMNSAL 180
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      264 GRSPMLVGNELTVADVLMVLTQGTGSSGAAPTNNQRMKSCENLAPF 312
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 GKSPMLVGNELTVADVLMVLTQGTGSSGAAPTNNQRMKSCENLAPF 229
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

```

US-10-463-676-5
; Sequence 5, Application US/10463676
; Publication No. US20040175375A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Sunghoon
; APPLICANT: Park, Bum-Joon

```

```

; TITLE OF INVENTION: Method for Treating Cancer Using P38/JTV-1 and Method
; FILE REFERENCE: 012679-091
; CURRENT APPLICATION NUMBER: US/10/463,676
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: KR 10-2003-13058
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 161
; TYPE: PRP
; ORGANISM: Homo sapiens
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(161)
; OTHER INFORMATION: 1-161 amino acid sequence of p38/JTV-1
US-10-463-676-5

```

```

Query Match      44.4%; Score 735; DB 16; Length 161;
Best Local Similarity 88.8%; Pred. No. 2.5e-66;
Matches 143; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

```

```

Qy      1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPATDAGHYOETSEPSILQALSROD 60
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MPMYQVRYHGGAPLRLVELPTCMYRLPNVHSKTTSPATDAGHYOETSEPSILQALSROD 60
Qy      61 DILRLVELKAAVDGSKMQLTPDADLVNIIQADEPTTLTNALDLSVLGKDYALK 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 DILRLVELKAAVDGSKMQLTPDADLVNIIQADEPTTLTNALDLSVLGKDYALK 120
Qy      121 DIVINANPASPPLSLVLRHLRCERYRLSTVHTSSVKVNPENLVYCFEGQARKSRHE 161
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 DIVINANPASPPLSLVLRHLRCERYRLSTVHTSSVKVNPENLVYCFEGQARKSRHE 161

```

RESULT 4

```

US-10-437-963-126132
; Sequence 126132, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; PRIOR FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 126132
; LENGTH: 925
; TYPE: PRP
; ORGANISM: Oryza sativa
; NAME/KEY: Oryza sativa
; OTHER INFORMATION: Clone ID: PAT_MRT4530_28709C.1.pcp
US-10-437-963-126132

```

```

Query Match      6.4%; Score 105.5; DB 16; Length 925;
Best Local Similarity 20.3%; Pred. No. 0.51;
Matches 64; Conservative 46; Mismatches 107; Indels 99; Gaps 13;

```

```

Qy      25 YRLPNVHSKTTSPATDAGHYOETSEPSILQALSROD--LKRRLVELKAAVDGSKMHT 82
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      152 YWVDNIIAKVGPV-----TIDPRLOALYKKTTELVGKSEKLVKLTSLGDDVHA 203
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      83 PDADLVNIIQAD--EPTTLTNALDLSVLGKDYA-----LKDIVINAN 127
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: February 23, 2005, 13:34:50 / Search time 13.3213 Seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-6  
Perfect score: 1655  
Sequence: 1 MPWGVKPKHYGSGAPRLVRL.....RWLKSCEMLAPSTALQLLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126.5	7.6	719	2 T52043	probable glutamate
2	112	6.8	728	2 T01200	probable glutamate
3	104	6.3	980	2 T30089	probable zinc prot
4	99	6.0	691	2 B89797	glycerol ester hyd
5	98.5	6.0	222	2 G82441	probable glucathio
6	98.5	6.0	377	2 F69172	hypothetical prote
7	98	5.9	690	2 A24545	triacylglycerol li
8	96.5	5.8	1440	1 STRUQT	multifunctional am
9	95.5	5.8	753	2 T24869	hypothetical prote
10	95	5.7	488	1 H64313	corrinoid/iron-sul
11	95	5.7	552	2 D82421	conserved hypotet
12	95	5.7	2655	2 D96595	probable acetyl-co
13	94.5	5.7	2471	2 T42977	large tegument pro
14	94	5.7	435	2 T01567	hypothetical prote
15	94	5.7	436	2 C86719	gtp-binding protei
16	93.5	5.6	1064	2 B86465	probable protein k
17	93	5.6	490	2 B84999	ketol-acid reducto
18	92	5.6	754	2 S37403	transcription fact
19	91.5	5.5	436	2 T51237	translacion elonga
20	91	5.5	1265	2 T47626	structural mainten
21	90.5	5.5	359	2 T37921	ras-associated pro
22	90.5	5.5	456	2 C84764	hypothetical prote
23	90.5	5.5	642	2 T39490	transcriptional int
24	90.5	5.5	716	2 T37830	probable glutamate
25	90	5.4	437	2 T51238	translacion elonga
26	90	5.4	559	2 A38604	poly(3-hydroxyalka
27	90	5.4	2954	2 T14156	kinesin-related pr
28	89.5	5.4	317	2 T05528	hypothetical prote
29	89.5	5.4	1597	2 S65053	genome polypeptide

30	89.5	5.4	1714	1 S18644	multifunctional am
31	89.5	5.4	4976	2 T14165	peptide synthetase
32	89	5.4	184	2 S74356	glutathione S-tran
33	89	5.4	763	2 AE2443	penicillin-binding
34	89	5.4	1023	2 T3169	neural zinc finger
35	89	5.4	1056	1 G02157	kinesin-like spind
36	89	5.4	1247	2 A33812	interphotoreceptor
37	88.5	5.3	602	2 AB3542	gtp-binding protei
38	88.5	5.3	656	2 H84206	acetyl-CoA synthet
39	88	5.3	821	2 T24728	hypothetical prote
40	88	5.3	2279	2 T42531	acetyl-CoA carboxy
41	88	5.3	2280	2 T38906	acetyl-CoA carboxy
42	87.5	5.3	474	2 B70597	hypothetical prote
43	87.5	5.3	820	2 H82302	ATP-dependent heli
44	87	5.3	570	2 T15763	hypothetical prote
45	86	5.2	906	2 T45158	pre-mRNA splicing

## ALIGNMENTS

RESULT 1  
T52043  
probable glutamate-tRNA ligase (EC 6.1.1.17) [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004  
C:Accession: T52043  
R:Day, I.S.; Golovkin, M.; Reddy, A.S.  
Biochim. Biophys. Acta 1399, 219-24, 1998  
A:Title: Cloning of the cDNA for glutamy1-tRNA synthetase from Arabidopsis thaliana.  
A:Reference number: 224836; PMID:9765600; PMID:9765600  
A:Accession: T52043  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-719 <DAY>  
A:Cross-references: UNIPROT:O82462; EMBL:AF067773; PDB:1AC36469.1  
C:Superfamily: Yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology  
C:Keywords: ligase

Query Match 7.6%; Score 126.5; DB 2; Length 719;  
Best Local Similarity 28.0%; Pred. No. 0.017;  
Matches 49; Conservative 22; Mismatches 57; Indels 47; Gaps 8;

Qy	128	PASPLSLVLRLLCEYRVLTSTHTSSVKNVPELVKCRGEQAKRSRREYQLGFL	187
Db	10	PESPLSVLVALSLSPV---TIDSSAAATTVPSFVF---SPGRKLN-----GATV	55
Qy	188	IKNVNPKTQMKFSVQTMCPIDEGENIARFLFSLFGCKHNAVTLTIDSVVDIAMFQLREG	247
Db	56	LIRVY-----GRSAKKLPDFVG--NNAPDSQIDENVDAVSV-FSSG	94
Qy	248	SKKRAVFRSMNSALGRSPMLVGNELTVADVVLVSLQOTGSGGAAPTQVRW	302
Db	95	SEFENAC--GRVDKLESSTFLVGHSLIADVAIWSALAGTS-----QRW	137

RESULT 2  
T01200  
probable glutamate-tRNA ligase (EC 6.1.1.17) P21E10.12 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
R:Davidson, S.; Rohlfing, T.; David, M.; O'Brian, D.  
Submitted to the EMBL Data Library, April 1999  
A:Description: The sequence of A. thaliana P21E10.  
A:Reference number: 214258  
A:Accession: T01200  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-728 <DAY>  
A:Cross-references: UNIPROT:O65253; EMBL:AF058914; NID:G3047074; PID:G3047084; GSPDB:GNO  
C:Experimental source: cultivar Columbia

A:Gene: ATSP:F21E10.12  
A:Map position: 5  
A:Introns: 47/2; 89/3; 141/1; 503/3; 659/3  
C:Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
F:223-499/Domain: glutamine-tRNA ligase homology <EGL>

Query Match 6.8%; Score 112; DB 2; Length 728;  
Best Local Similarity 26.6%; Pred. No. 0.29; Mismatches 57; Indels 56; Gaps 9;  
Matches 49; Conservative 22; Mismatches 57; Indels 56; Gaps 9;

QY 128 PASPPSLVLRLLCERYRLSTVHTSSVKNVPELVKCFEQARKQSRHEYOGLFTL 187  
DB 10 PESPPSLVYVLSASPV---TIDSSAAATVPSPFV---SDGRKLN-----GATV 55  
QY 188 IMKNVPTQMKFSVQTMCPRIEGBGNVAFLEFLSGQKHNATLT-----LIDSWD 238  
DB 56 LKRYV-----GRSAKKLPDFYG--NNAFDSQVSIILCINMKIDWVD 95  
QY 239 IAMEFLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSLVQQTGSSGAAPTN 298  
DB 96 YASV-FSSGSEFENAC--GRVDKLTESSTFLVGHSLIDVVAIWSALAGTG----- 143  
QY 299 VQRW 302  
DB 144 -QRW 146

RESULT 3  
T30089  
Probable zinc proteinase (EC 3.4.24.-) C02G6.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30089  
R:Bentley, D.; Kemp, K.; Scheer, P.  
Submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid C02G6.  
A:Reference number: Z20734  
A:Accession: T30089  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-980 <BEN>  
A:Cross-references: UNIPROT:Q17592; EMBL:U55372; PDB:AAA98001.1; GSPDB:GN00023; CESP:CO  
A:Experimental source: strain Bristol N2; clone C02G6  
C:Genetics:  
A:Gene: CESP:C02G6.1  
A:Map position: 5  
A:Introns: 25/3; 215/2; 266/3; 540/3; 585/3; 786/1; 898/2  
C:Superfamily: insulin-degrading enzyme (IDE)  
C:Keywords: hydrolase; metalloproteinase; zinc  
F:70/74/Binding site: zinc (His) #status predicted  
F:73/Active site: Glu #status predicted

Query Match 6.3%; Score 104; DB 2; Length 980;  
Best Local Similarity 20.1%; Pred. No. 2.1; Indels 142; Gaps 22;  
Matches 86; Conservative 67; Mismatches 133; Indels 142; Gaps 22;  
QY 10 HGGAPRLVELPTCMYRLPNVHSTKTPATDAGHVOETSEPSLQALERSODILKRLYE- 68  
DB 276 HNGPSLVELKRLGM-VNSLKSDENTIAAGFGLINVTMDLSTGLLEN-VDEIIQMLNLY 333  
QY 69 ---LKAADVGLSKMHTDADI-DVT-NILOADEPTLATNT-----LIDNSVLGKDY- 116  
DB 334 IGMKLSF--GPGQWHDLDLADSDVKFRPKQKQEMKAINIATSLQYIPIEHLSSRYL 391  
QY 117 -----GALKDQIVINANPAPPLSLVLRLLCERYR----- 147  
DB 392 LTKYPERIKELSLTLPSTN-----MLVIVSQKFKQEGEENTNEPVYTEMKTDLSPE 445  
QY 148 -----VLSTVH--THSSVKNVPELVKCFEQARKQSRHEYOGL-----GFTLIW----- 189  
DB 446 KMKYENALKTSHALHLPKRN--EYIATNFGQKPRRSVKNHFKLISDDGWSRWFQKD 503

QY 190 --KNVEKTMKFSVQF-----MCPKEG-----GNIAFLPSL 220  
DB 504 DEYNMFKQSTKFTALTPIVSGNPRISLISLMLMCFCDILSETYNALAGICQFELSP 563  
QY 221 FG-----QGNNAVTL-----TLIDSWVDIAMPQRESSKE 251  
DB 564 FGVQKQSTQDREAERHASTLHVYGDQKPLFVGHILTSICMIFKIDRTREFVLFESIK- 622  
QY 252 KAVERSMNSALGRSPMLVG--NELTVADVLMVSLVQQTGSSGAAPTNVQRMKSGEN 308  
DB 623 -----RTLINNAFSPYLLTQHYNQLIVDKV-MSKEQLLAVCDSTLENVQGFAR--EM 674  
QY 309 IAPFSTAL 316  
DB 675 IQAFHMEI 682

RESULT 4  
B89797  
glycerol ester hydrolase [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: B89797  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; PMID:21311952; PMID:11418146  
A:Accession: B89797  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-691 <KUR>  
A:Cross-references: UNIPROT:Q99W06; GB:BA000018; PID:913700235; PDB:BAB41533.1; GSPDB:G  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: geh  
C:Superfamily: *Staphylococcus triacylglycerol lipase*

Query Match 6.0%; Score 99; DB 2; Length 691;  
Best Local Similarity 20.2%; Pred. No. 3.4;  
Matches 65; Conservative 53; Mismatches 110; Indels 94; Gaps 15;  
QY 1 MPWQVPRVYHGGAPRLVELPTCMYRLPNVHSTKTPATDA-----GHVOETSEPSL 52  
DB 172 IPQKVLPRNDKAAPISTTPS-----NDKTAKSTYADATTDKHPNQDTHQPAH 223  
QY 53 QALERSODILKRLYEIKAVDGLSKMI--HTPDALD-----VTNIIQADEPTTL 101  
DB 224 QIIDAKQDDTV--RQSQKPVQVGLSKHIDGQNSPEKPTDKNTDNQILIDALQA--PKTR 280  
QY 102 AATVLDLNSVLTGQVQALDKDQIVINANPAPPLSLVLRLLCERYRLSTVHTSSVKNV 161  
DB 281 STTNMAAD-----AKVRPLKANQVQF-----LNKYPV--VFVHGFGLIV 318  
QY 162 PENLVKCFG-----EQARKQSRHEYOGLFTLWKNVPK-TQMKFSVQTMCPKEG 209  
DB 319 GDAPRLVINYWGNGKFKYIEIRKQGVNVHQAASAFSGSNVDRVVELTYTKGRGVYD 378  
QY 210 EGNIAFLPSLFGQKHNAV-----TLTIDSWVDIAMPQRESSS 249  
DB 379 AAHDAKYGHERYQKTYKGIIMPWEPQKVHIVGHSMGGQITLMBEF-----LRNG-N 430  
QY 250 KEKAAVFRSMNSALGRSPMLVG 271  
DB 431 KEELATYHKAHGERT--SPLFTG 450

RESULT 5  
G82441  
probable glutathione S-transferase VCA0584 [imported] - *Vibrio cholerae* (strain N16961 s  
C:Species: *Vibrio cholerae*  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:33:04 ; Search time 60.8145 Seconds  
(without alignments)  
2694.512 Million cell updates/sec

Title: US-10-622-817-6  
Perfect score: 1655  
Sequence: 1 MPWQVQKPRYHGSAFLRVRL.....RWLKSCEMLAPFTALQLLK 320

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1655	100.0	320	Q8R010	Q8R010 mus musculu
2	1652	99.8	320	Q8R2Y6	Q8R2Y6 mus musculu
3	1573	95.0	320	1 MCA2_CRIGR	Q9WY77 cricetulus
4	1464	88.5	320	1 MCA2_HUMAN	Q13155 homo sapien
5	1460	88.2	320	2 Q96CZ5	Q96CZ5 homo sapien
6	1412	85.3	280	2 Q8R3V2	Q8R3V2 mus musculu
7	1153.5	69.7	311	2 Q6DK86	Q6DK86 xenopus tro
8	1146.5	69.3	311	2 Q6TNU4	Q6TNU4 xenopus lae
9	1146.5	69.3	311	2 Q7ZXD7	Q7ZXD7 xenopus lae
10	926.5	56.0	321	2 Q773C0	Q773C0 brachydanio
11	247	14.9	340	2 Q7Q7A3	Q7Q7A3 anopheles g
12	224	13.5	301	2 Q7KUM5	Q7KUM5 drosophila
13	224	13.5	313	2 Q8T060	Q8T060 drosophila
14	224	13.5	322	2 Q6NKM4	Q6NKM4 drosophila
15	216.5	13.1	334	1 MCA2_DROME	Q9VUR3 drosophila
16	126.5	7.6	719	2 Q82462	Q82462 arabidopsis
17	112	6.8	728	2 Q65253	Q65253 arabidopsis
18	107	6.5	715	2 Q9L1Z8	Q9L1Z8 oryza sativ
19	106	6.4	328	2 Q86X73	Q86X73 homo sapien
20	106	6.4	559	2 Q8RPZ6	Q8RPZ6 pseudomons
21	106	6.4	869	2 Q6PD57	Q6PD57 homo sapien
22	106	6.4	1721	2 Q7NID9	Q7NID9 gloeobacter
23	104	6.3	980	2 Q17592	Q17592 caenorhabdi
24	102.5	6.2	420	2 Q8RFOO	Q8RFOO corynebacte
25	102	6.2	559	2 Q939A8	Q939A8 pseudomons
26	102	6.2	1048	2 Q7N3E2	Q7N3E2 photoxabanu
27	101.5	6.1	715	2 Q91B77	Q91B77 white spot
28	101.5	6.1	730	2 Q910F5	Q910F5 white spot
29	101.5	6.1	913	2 Q6DRB3	Q6DRB3 brachydanio
30	101.5	6.1	913	2 Q6PFO4	Q6PFO4 brachydanio
31	100.5	6.1	555	2 Q7MEV7	Q7MEV7 vibrio vuln

32	100.5	6.1	555	2	Q8D7U3	Q8D7U3 vibrio vuln
33	100.5	6.1	922	2	Q8K2J4	Q8K2J4 mus musculu
34	99.5	6.0	885	2	Q6REMS	Q6REMS brachydanio
35	99.5	6.0	559	2	Q8RQ67	Q8RQ67 pseudomons
36	99	6.0	690	1	LIP_STRAW	LIP_STRAW
37	99	6.0	690	2	Q6GCF1	Q6GCF1 staphylococ
38	99	6.0	691	2	Q798Z7	Q798Z7 staphylococ
39	99	6.0	691	2	Q9WQ06	Q9WQ06 staphylococ
40	99	6.0	691	2	Q7A7P2	Q7A7P2 staphylococ
41	98.5	6.0	222	2	Q9KM05	Q9KM05 vibrio chol
42	98.5	6.0	377	2	Q26161	Q26161 mechanobac
43	98.5	6.0	1597	2	Q8V9A2	Q8V9A2 crucifer to
44	98	5.9	240	2	Q8BTR1	Q8BTR1 mus musculu
45	98	5.9	690	1	LIP_STRAW	LIP_STRAW

## ALIGNMENTS

RESULT 1  
Q8R010 PRELIMINARY; PRT: 320 AA.  
ID Q8R010;  
AC Q8R010;  
DT 01-JUN-2002 (TRENBLREL. 21, Created)  
DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)  
DT 05-JUL-2004 (TRENBLREL. 27, Last annotation update)  
DE Utrv-pending protein.  
GN Name=Utrv;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N, and FVB/N-3; TISSUE=Mammary tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Boask S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny K.C., Harte S., Sodergren B.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N-3; TISSUE=Mammary tumor;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC024410; AAH24410.1; -;  
DR EMBL; BC026972; AAH26972.1; -;  
DR MGD; MGI:2385237; Utrv.  
DR InterPro; IPR004046; GST\_C-term.  
DR InterPro; IPR010987; GST\_C-like.  
DR Pfam; PF00043; GST\_C\_1; -;  
SQ SEQUENCE 320 AA; 35396 MW; 1745D7BEABCB3670D CRC64;

Query Match 100.0%; Score 1655; DB 2; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 2, 6e-128;  
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYVKKYHGGSAFLRVELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOLESROD 60  
 1 MPWYVKKYHGGSAFLRVELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOLESROD 60  
 DB 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 DB 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 QY 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300  
 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300  
 DB 301 RWLKSCEMLAPFSTALQULK 320  
 301 RWLKSCEMLAPFSTALQULK 320

QY 301 RWLKSCEMLAPFSTALQULK 320  
 301 RWLKSCEMLAPFSTALQULK 320

DB 301 RWLKSCEMLAPFSTALQULK 320  
 301 RWLKSCEMLAPFSTALQULK 320

RESULT 2  
 Q8R2Y6 PRELIMINARY; PRT; 320 AA.  
 ID Q8R2Y6;  
 AC Q8R2Y6;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Jtlv1-pending protein.  
 GN Name=Jtlv1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECB II; TISSUE=Mammary tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stachleon M., Soares M.B., Bonaldo M.F., Rubin G.M., Hong L.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,  
 Boeck S.A., McBwan P.J., McKernan K.J., Matk J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley A.C., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Skalski U., Schmitz J., Myers R.M., Butcherfield Y.S.,  
 Krzywinski M.I., Skalski U., Small D.E., Scherch A., Schein J.E.,  
 Jones S.J., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECB II; TISSUE=Mammary tumor;  
 RX Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 RA EMBL; BC026958; AAH26958.1; -.

DR MGI:2385237; Jtlv1.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR010987; GST\_C\_1like.  
 DR Pfam; PF00043; GST\_C\_1.  
 SQ SEQUENCE 320 AA; 35423 MW; 1C21F1A74C9882B4 CRC64;

Query Match 99.8%; Score 1652; DB 2; Length 320;  
 Best Local Similarity 99.7%; Pred. No. 4, 6e-128;  
 Matches 319; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPWYVKKYHGGSAFLRVELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOLESROD 60  
 1 MPWYVKKYHGGSAFLRVELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOLESROD 60  
 DB 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 QY 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 DB 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 61 DILKRLYEKAADVGLSKMHTPDADVTNIIQADEPTLATNTLDLSVGLGYALK 120  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 DB 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVHTHSSVKVNPENLVKCFGEQARKOSRHE 180  
 QY 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 DB 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 181 YOLGFTLIWKNVPTQMKFSVQTMCPLEGEGNIARFLFSIFGQKNAVTLTLDISWDIA 240  
 QY 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300  
 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300  
 DB 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300  
 241 MFOLREGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSVLQOTGSSGAAPTNYQ 300

QY 301 RWLKSCEMLAPFSTALQULK 320  
 301 RWLKSCEMLAPFSTALQULK 320

DB 301 RWLKSCEMLAPFSTALQULK 320  
 301 RWLKSCEMLAPFSTALQULK 320

RESULT 3  
 MCA2\_CRIGR STANDARD; PRT; 320 AA.  
 ID MCA2\_CRIGR  
 AC Q9WVW7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Multisynthetase complex auxiliary component p38.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 NC NCB1\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=9906915; PubMed=9878398; DOI=10.1006/jmbi.1998.2316;  
 RA Quevillon S., Robinson J.-C., Berthomieu E., Stalicka M., Mirande M.,  
 "Macromolecular assembly of aminoacyl-tRNA synthetases:  
 RT identification of protein-protein interactions and characterization of  
 a core protein.";  
 RL J. Mol. Biol. 285:183-195(1999).  
 CC -1- FUNCTION: Probable core protein of the multisynthetase complex  
 that serves as a template for the assembly of the supramolecular  
 structure.  
 CC -1- SUBUNIT: Component of the multisynthetase complex which is  
 CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the  
 CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,  
 CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary  
 CC proteins, p16, p48 and p43.  
 CC -1- SIMILARITY: Contains 1 GST-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF072727; AAD34423.1; -.
CC InterPro: IPR010987; GST_C_1like.
CC InterPro: IPR004046; GST_Cterm.
CC Pfam: PF00043; GST_C.1.
CC Protein biosynthesis.
CC SEQUENCE 320 AA; 35433 MW; 6D24E033ABEC610A CRC64;

Query Match 95.0%; Score 1573; DB 1; Length 320;
Best Local Similarity 94.4%; Pred. No. 1.6e-121;
Matches 302; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MPMYGVKPYHGSGAPLRVELPTCYRLPNVHSKTTSPATDAGHVOETSPSIOALESROD 60
Db 1 MPMYGVKPYHGSGAPLRVELPTCYRLPNVHSKTTSPATDAGHVOETSPSIOALESROD 60
Qy 61 DILKRLVELKAAVDSLKSMHTPDADLVNTILQADEPTLTATNTLDINSVLGKDYGLK 120
Db 61 DILKRLVELKAAVDSLKSMHTPDADLVNTILQADEPTLTATNTLDINSVLGKDYGLK 120
Qy 121 DIVINANASPLSLVLRHLICERYRLSTVHTSSVKNVPELVICFGEQARKQSRHE 180
Db 121 DIVINANASPLSLVLRHLICERYRLSTVHTSSVKNVPELVICFGEQARKQSRHE 180
Qy 181 YOLGFTLIMKNVPTQMKFSVQTMCPIGEGNIARFLPSLFGCKHNVTLTLIDSWVIA 240
Db 181 YOLGFTLIMKNVPTQMKFSVQTMCPIGEGNIARFLPSLFGCKHNVTLTLIDSWVIA 240
Qy 241 MFOLEGGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGAAPTVOQ 300
Db 241 MFOLEGGSSKEKAAVFRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGAAPTVOQ 300
Qy 301 RWLKSCEMLAPSTMLQILK 320
Db 301 RWLKSCEMLAPSTMLQILK 320
Qy 301 RWLKSCEMLAPSTMLQILK 320
Db 301 RWLKSCEMLAPSTMLQILK 320

RESULT 4
MCA2_HUMAN STANDARD; PRT; 320 AA.
ID MCA2_HUMAN
AC Q13155; Q9P1L2;
DT 01-NOV-1997 (Rel. 35; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 05-JUL-2004 (Rel. 44; Last annotation update)
DE Multisynthetase complex auxiliary component p38 (JTV-1 protein)
DE (PRO00992).
GN Name=JTV1;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Hmo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96115582; PubMed=8666379;
RA Nicolaides N.C., Kinzler K.W., Vogelstein B.;
RT "Analysis of the 5' region of PMS2 reveals heterogeneous transcripts
RT and a novel overlapping gene.";
RL Genomics 29:329-334(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=2338857; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Straube R.G., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marutka K., Farmer A.A., Rubin J., Hong L.,
RA Stepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedl T.B., Toshiyuki S., Carinuci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnarone P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W.,

```

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 197-320 FROM N.A.
RC TISSUE=Fetal Liver;
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 121 new genes
RT deduced by analysis of cDNA clones from human fetal liver.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP INTERACTION WITH FUBP1.
RX MEDLINE=22716800; PubMed=12819782; DOI=10.1038/ng1182;
RA Kim M.J., Park B.-J., Kang Y.-S., Kim H.-J., Park J.-H., Kang J.W.,
RA Lee S.W., Han J.M., Lee H.-W., Kim S.;
RT "Downregulation of FUBS-binding protein and c-myc by RNA synthetase
RT cofactor p38 is required for lung cell differentiation.";
RL Nat. Genet. 34:330-336(2003).
CC -1- FUNCTION: Probable core protein of the multisynthetase complex
CC that serves as a template for the assembly of the supramolecular
CC structure. Mediates ubiquitination of FUBP1 and its degradation by
CC the proteasome.
CC -1- SUBUNIT: Component of the multisynthetase complex which is
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, lysyl,
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary
CC proteins, p18, p48 and p43. Binds FUBP1.
CC -1- SIMILARITY: Contains 1 GST-like domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 312.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U24169; AAC50391.1; ALT FRAME.
CC EMBL: BC002853; AAH02853.1; -.
CC EMBL: BC010156; AAH010156.1; -.
CC EMBL: AF116615; AAF71039.1; -.
CC H-InvDB: HIX006460; -.
CC MIM: 600859; -.
CC InterPro: IPR010987; GST_C_1like.
CC InterPro: IPR004046; GST_Cterm.
CC Pfam: PF00043; GST_C.1.
CC Protein biosynthesis.
CC SEQUENCE 320 AA; 35349 MW; F253726B63C12BAB CRC64;

Query Match 88.5%; Score 1464; DB 1; Length 320;
Best Local Similarity 87.2%; Pred. No. 1.6e-112;
Matches 279; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MPMYGVKPYHGSGAPLRVELPTCYRLPNVHSKTTSPATDAGHVOETSPSIOALESROD 60
Db 1 MPMYGVKPYHGSGAPLRVELPTCYRLPNVHSKTTSPATDAGHVOETSPSIOALESROD 60
Qy 61 DILKRLVELKAAVDSLKSMHTPDADLVNTILQADEPTLTATNTLDINSVLGKDYGLK 120
Db 61 DILKRLVELKAAVDSLKSMHTPDADLVNTILQADEPTLTATNTLDINSVLGKDYGLK 120
Qy 121 DIVINANASPLSLVLRHLICERYRLSTVHTSSVKNVPELVICFGEQARKQSRHE 180
Db 121 DIVINANASPLSLVLRHLICERYRLSTVHTSSVKNVPELVICFGEQARKQSRHE 180

```

```

Db      121 DIVINAMPASPPSLVLVLRLLCEHFRVLTSTVHTSSVKSPENLLKCFGEONKKOPROD 180
Qy      181 YOLGFTLLWKVVPKTKOMKFSVQTMCPIDEGENIARFLPSLFGQKNAVTLTIDSWDIA 240
Db      181 YOLGFTLLWKVVPKTKOMKFSIQTMCPIDEGENIARFLPSLFGQKNAVNTLTIDSWDIA 240
Qy      241 MFOLREGSSKKAFAVRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTMVQ 300
Db      241 IFOLKEGSSKKAFAVRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTMVQ 300
Qy      301 RMLKSCENIAPFPSTALQULK 320
Db      301 RWMRSCENIAPFPNTALKLX 320

```

```

RESULT 5
096CZ5 PRELIMINARY; PRT; 320 AA.
ID 096CZ5 AC 096CZ5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE JTV1.
GN Name=JTV1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schlier G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abrahams R.D., Mullaly S.J.,
RA Richards S., Worley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013630; AAH13630.1; -.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C-like.
DR Pfam; PF00043; GST_C_1_-like.
SQ SEQUENCE 320 AA; 35335 MW; 19F14BF58612E08 CRC64;

```

```

Query Match      88.2%; Score 1460; DB 2; Length 320;
Best Local Similarity 86.9%; Pred. No. 3.4e-112;
Matches 278; Conservative 16; Mismatches 26; Indels 0; Gaps 0;
Qy      1 MPMYGVKHYHGASAPLRYELPTCYVRLPNVHSKTTPSPATDAGVHOENSPSTOLAESROD 60
Db      1 MPMYGVKHYHGASAPLRYELPTCYVRLPNVHSGRSTGPPGAGHVOESNLSLQLESROD 60
Qy      61 DILKRLYEIKAAVDGLSKMHTPPADLDVTNIIQADEPTTLATNTLDLSVGLKDYGALK 120

```

```

Db      61 DILKRLYEIKAAVDGLSKMHTPPADLDVTNIIQADEPTTLATNTLDLSVGLKDYGALK 120
Qy      121 DIVINAMPASPPSLVLVLRLLCEHFRVLTSTVHTSSVKSPENLLKCFGEONKKOPROD 180
Db      121 DIVINAMPASPPSLVLVLRLLCEHFRVLTSTVHTSSVKSPENLLKCFGEONKKOPROD 180
Qy      181 YOLGFTLLWKVVPKTKOMKFSVQTMCPIDEGENIARFLPSLFGQKNAVTLTIDSWDIA 240
Db      181 YOLGFTLLWKVVPKTKOMKFSIQTMCPIDEGENIARFLPSLFGQKNAVNTLTIDSWDIA 240
Qy      241 MFOLREGSSKKAFAVRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTMVQ 300
Db      241 IFOLKEGSSKKAFAVRSMNSALGRSPMLVGNELTVADVLMVSLQOTGSSGGAAPTMVQ 300
Qy      301 RMLKSCENIAPFPSTALQULK 320
Db      301 RWMRSCENIAPFPNTALKLX 320

```

```

RESULT 6
0983V2 PRELIMINARY; PRT; 280 AA.
ID 0983V2 AC 0983V2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE JTV1.
GN Name=JTV1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schlier G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abrahams R.D., Mullaly S.J.,
RA Richards S., Worley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024480; AAH24480.1; -.
DR MGD; MGI:2385237; JTV1.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C-like.
DR Pfam; PF00043; GST_C_1_-like.
SQ SEQUENCE 280 AA; 31097 MW; E5BB81498983FD2E CRC64;

```

```

Query Match      85.3%; Score 1412; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 2.6e-108;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      45 QETSPRSQALAESRODDILKRLYEIKAAVDGLSKMHTPPADLDVTNIIQADEPTTLATN 104

```

```
Db 5 OETSPEISLOALESRODDILKRLYELKAAVDGSKMHTFDADLDVTNLIQADEPTLATN 64
Qy 105 TLDLNSVYGVKQYKALKDVTYNNPASPISLIVLHRLICERYVLTSTVHTSSVKNVPPN 164
Db 65 TLDLNSVYGVKQYKALKDVTYNNPASPISLIVLHRLICERYVLTSTVHTSSVKNVPPN 124
Qy 165 LVKCGEQARKOSRHEHYQLFTLIMKNVPTQMKFSVQTMCPIDEGENIARFLFSLFGOK 224
Db 125 LVKCGEQARKOSRHEHYQLFTLIMKNVPTQMKFSVQTMCPIDEGENIARFLFSLFGOK 184
Qy 225 HNAVTLTLLIDSVWDIAMFOLREGSSKEKAIVFRSMNSALGRSPMLVGNELTVADVYLMGV 284
Db 185 HNAVTLTLLIDSVWDIAMFOLREGSSKEKAIVFRSMNSALGRSPMLVGNELTVADVYLMGV 244
Qy 285 LQQTGSSGAAPTNNQRMKSCENLAPFSTALQLLK 320
Db 245 LQQTGSSGAAPTNNQRMKSCENLAPFSTALQLLK 280
```

## RESULT 7

```
Qy 06DK86 PRELIMINARY; PRT; 311 AA.
AC 06DK86;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE MGC69221 protein.
GN Name=MGC69221;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8364;
```

```
KN [1]
RP TISSUE=Embryo;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Pelngold E.A., Grove L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshynki S., Carrinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP TISSUE=Embryo;
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL, BC074561; AAH74561.1;
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1_1like.
DR SEQUENCE 311 AA; 3480 MW; D98P27F73C466154 CRC64;
```

Query Match 69.7%; Score 1153.5; DB 2; Length 311;  
Best Local Similarity 67.5%; Pred. No. 6.8e-87;  
Matches 216; Conservative 49; Mismatches 46; Indels 9; Gaps 3;

```
Qy 1 MEMYQVKYHGGSABRLAYELPTCMYRLPNVHSKTSIPATDAGHVOETSPESLOALESROD 60
Db 1 MEMYQVKYHGGSABRLAYELPTCMYRLPNVHSKTSIPATDAGHVOETSPESLOALESROD 52
Qy 61 DILKRLYELKAAVNDGSKMHTFDADLDVTNLIQADEPTLATNTLDLNSVYGVKQYKAL 120
Db 53 DILKRLYELKAAVNDGSKMHTFDADLDVTNLIQADEPTLATNTLDLNSVYGVKQYKAL 112
Qy 121 DIVINANPASPISLIVLHRLICERYVLTSTVHTSSVKNVPPN LVKCGEQARKOSRHE 180
Db 113 DIVINANPASPISLIVLHRLICERYVLTSTVHTSSVKNVPPN LVKCGEQARKOSRHE 172
Qy 181 YQLGFTLIMKNVPTQMKFSVQTMCPIDEGENIARFLFSLFGOKHNAVTLTLLIDSVWDIA 240
Db 173 YQLGFTLIMKNVPTQMKFSVQTMCPIDEGENIARFLFSLFGOKHNAVTLTLLIDSVWDIA 232
Qy 241 MFOLREGSSKEKAIVFRSMNSALGRSPMLVGNELTVADVYLMGV LQQTGSSGAAPTNNQ 300
Db 233 MFOLREGSSKEKAIVFRSMNSALGRSPMLVGNELTVADVYLMGV LQQTGSSGAAPTNNQ 291
Qy 301 PMLKSCENLAPFSTALQLLK 320
Db 292 PMLKSCENLAPFSTALQLLK 311
```

## RESULT 8

```
Qy 06IND4 PRELIMINARY; PRT; 311 AA.
AC 06IND4;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE MGC80304 protein.
GN Name=MGC80304;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
```

```
KN [1]
RP TISSUE=Ovary;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Pelngold E.A., Grove L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshynki S., Carrinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP TISSUE=Ovary;
RC TISSUE=Ovary;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RL Submitted (JUN-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL, BC074561; AAH74561.1;
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C_1like.
DR Pfam; PF00043; GST_C_1_1like.
DR SEQUENCE 311 AA; 3480 MW; D98P27F73C466154 CRC64;
```

RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC072178; AAH72178.1; -  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR010987; GST\_C\_1like.  
 DR Pfam; PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34424 MW; D5E8325C18D88751 CRC64;

Query Match  
 Best Local Similarity 69.3%; Score 1146.5; DB 2; Length 311;  
 Matches 218; Conservative 44; Mismatches 49; Indels 9; Gaps 3;

QY 1 MPMTQVKEPHGSGAPLRYELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOALERSOD 60  
 DB 1 MPMTQVKEPHGSGA--IQVDLPTCMYRLPNVSKTTPATDAGHVOETSEPSLOALERSOD 52  
 QY 61 DILKRLVYLKAAVGLSKMTHTPDADLVNIIQADEPTLATNTLDLNSVYLGKYGALK 120  
 DB 53 DILKRLVYLKAAVGLSKMTHTPDADLVNIIQADEPTLATNTLDLNSVYLGKYGALK 112  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVTHSSVKNVPPNVLVKCFGEQARKOSRHE 180  
 DB 113 DIVINANPASPPLSLVLRHLRCERYVLTSTVTHSSVKNVPPNVLVKCFGEQARKOSRHE 172  
 QY 181 YOLGFTLIMKQVPTQMKFSVQTMCPTEGEGNTARFLFSIFGQKHNAVTLTLDLNSVWDA 240  
 DB 173 YOLGFTLIMKQVPTQMKFSVQTMCPTEGEGNTARFLFSIFGQKHNAVTLTLDLNSVWDA 232  
 QY 241 MPOLREGSSKEKAAPVRSNMSALGRSPMLVGNELTVADVVLMSVLYQGTGGSGAAPTVNQ 300  
 DB 233 IFOLREGSSKEKAAPVRSNMSALGRSPMLVGNELTVADVVLMSVLYQGTGGSGAAPTVNQ 291  
 QY 301 RMLKSCENLAPFSTALQILK 320  
 DB 292 KMKKSCENLAPFSTALQILK 311

## RESULT 9

ID Q72YD7 PRELIMINARY; PRT; 311 AA.  
 AC Q72YD7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Jv1-prov protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NC NCB1\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cavaant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Mair M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.,  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative.";  
 RL Dev. Dyn. 225:384-391 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.,  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC043832; AAH43832.1; -  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR010987; GST\_C\_1like.  
 DR Pfam; PF00043; GST\_C\_1.  
 SQ SEQUENCE 311 AA; 34647 MW; 2F08C9DA60DD63BE CRC64;

Query Match  
 Best Local Similarity 69.3%; Score 1146.5; DB 2; Length 311;  
 Matches 218; Conservative 44; Mismatches 49; Indels 9; Gaps 4;

QY 1 MPMTQVKEPHGSGAPLRYELPTCMYRLPNVSKTTPATDAGHVOETSEPSLOALERSOD 60  
 DB 1 MPMTQVKEPHGSGA--IQVDLPTCMYRLPNVSKTTPATDAGHVOETSEPSLOALERSOD 52  
 QY 61 DILKRLVYLKAAVGLSKMTHTPDADLVNIIQADEPTLATNTLDLNSVYLGKYGALK 120  
 DB 53 DILKRLVYLKAAVGLSKMTHTPDADLVNIIQADEPTLATNTLDLNSVYLGKYGALK 112  
 QY 121 DIVINANPASPPLSLVLRHLRCERYVLTSTVTHSSVKNVPPNVLVKCFGEQARKOSRHE 180  
 DB 113 DIVINANPASPPLSLVLRHLRCERYVLTSTVTHSSVKNVPPNVLVKCFGEQARKOSRHE 172  
 QY 181 YOLGFTLIMKQVPTQMKFSVQTMCPTEGEGNTARFLFSIFGQKHNAVTLTLDLNSVWDA 240  
 DB 173 YOLGFTLIMKQVPTQMKFSVQTMCPTEGEGNTARFLFSIFGQKHNAVTLTLDLNSVWDA 232  
 QY 241 MPOLREGSSKEKAAPVRSNMSALGRSPMLVGNELTVADVVLMSVLYQGTGGSGAAPTVNQ 300  
 DB 233 IFOLREGSSKEKAAPVRSNMSALGRSPMLVGNELTVADVVLMSVLYQGTGGSGAAPTVNQ 291  
 QY 301 RMLKSCENLAPFSTALQILK 320  
 DB 292 KMKKSCENLAPFSTALQILK 311

## RESULT 10

ID Q7T3CO PRELIMINARY; PRT; 321 AA.  
 AC Q7T3CO;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE zgc:63976.  
 GN ORFNames=zgc:63976;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NC NCB1\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,



RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel F.,  
RA Diachenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.T., Usdin T.B., Tohilyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Whillans S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Myers R.M., Butlerfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallegange D.E., Scherch J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Kidney;  
RL Strubeberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC053178; AAB53178.1; -  
DR ZFIN; ZDB-GENE-040426-2652; zgc:63976.  
DR InterPro; IPR004046; GST\_Cterm.  
DR InterPro; IPR010987; GST\_C\_1like.  
DR Pfam; PF00043; GST\_C\_1; 34852 MW; BAB69951208244A CRC64;  
SQ SEQUENCE 321 AA; 34852 MW; BAB69951208244A CRC64;  
  
Query Match 56.0%; Score 926.5; DB 2; Length 321;  
Best Local Similarity 57.3%; Pred. No. 4,1e-68;  
Matches 188; Conservative 48; Mismatches 77; Indels 15; Gaps 7;  
  
QY 1 MEMYQVYHGGASAPLVELPTCMYRLPNVHSKTTSP---PATDAGHVOETSEPSLOALES 57  
DB 1 MEMYQVYHGGASAPLVELPTCMYRLPNVHSKTTSP---PATDAGHVOETSEPSLOALES 54  
QY 58 RDDDLIKRLYEKAAVDGSKMHPDADLVNTI---LQADDEPTLATNTLDINSVLGK 114  
DB 55 RDDDLIKRLYEKAAVDGSKMHPDADLVNTI---LQADDEPTLATNTLDINSVLGK 114  
QY 115 DYGALKDVIYINANPASPLSLVLRHLICERYVSLVTHSSVKNVPELVKCFG-EQA 173  
DB 115 DYGALKDVIYINANPASPLSLVLRHLICERYVSLVTHSSVKNVPELVKCFG-EQA 173  
QY 174 KRQSRHEVYDGLTLLWKVNPXTQMKFSVQTCPIBEGENIARFLFELFG-QKHNAVTLTL 232  
DB 174 KRQSRHEVYDGLTLLWKVNPXTQMKFSVQTCPIBEGENIARFLFELFG-QKHNAVTLTL 232  
QY 175 HSYARHRFQLGFTLLIKVDVSKLQMKFSTQNMCPICEEGNVARFLYLLGAEPDPVSATL 234  
DB 175 HSYARHRFQLGFTLLIKVDVSKLQMKFSTQNMCPICEEGNVARFLYLLGAEPDPVSATL 234  
QY 233 IISWVDIANFOLREGSKKKAVERFMNALSFPVLVGNELTVANVLMVSLQQTGSSS 292  
DB 233 IISWVDIANFOLREGSKKKAVERFMNALSFPVLVGNELTVANVLMVSLQQTGSSS 292  
QY 235 MGVWVDIALFOLAEGGSKERRAANLRLNALGHSPLWLGQEFSLADIVASCCVLQGTGTS 294  
DB 235 MGVWVDIALFOLAEGGSKERRAANLRLNALGHSPLWLGQEFSLADIVASCCVLQGTGTS 294  
QY 293 GAAPTIVORWMLKSCENTLAPFTALQLLK 320  
DB 293 GAAPTIVORWMLKSCENTLAPFTALQLLK 320  
QY 295 -SAPANVQWMLKSCQNLGTFSCVDPLIQ 321  
DB 295 -SAPANVQWMLKSCQNLGTFSCVDPLIQ 321  
  
RESULT 11  
Q707A3 PRELIMINARY; PRT; 340 AA.  
AC Q707A3;  
DT 01-MAR-2004 (Tremblrel. 26, Created)  
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE AGCP5808 (Fragment).  
GN Name=agcgs0514; ORFName=ENGANG0000011827;  
OS Anopheles gambiae str. PEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
OC NCBI\_TaxID=180454;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=PEST;  
RA Anopheles Genome Sequencing Consortium;  
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAB801008960; EAA10870.1; -  
FT NON TER 1 1  
SQ SEQUENCE 340 AA; 37467 MW; FDD08525CD4EB976 CRC64;  
  
Query Match 14.9%; Score 247; DB 2; Length 340;  
Best Local Similarity 28.7%; Pred. No. 6,1e-12;  
Matches 94; Conservative 49; Mismatches 118; Indels 66; Gaps 13;  
  
QY 3 MYQVYHGGASAPLVELPTCMYRLPNVHSKTTSP-----ATD-----AGH----- 44  
DB 12 MYRLKPVW--NODVCELPCTCMYTLKPVCAVSHDPKASADSATGLAAGPVDPSIEID 69  
QY 45 ---QETSEPSLOALESRODDILKRLYEKAAVDGSKMHPDADLV--TNILQADDEPT 100  
DB 70 NLLQKAENDELMLMERQORVLQQLAELK-----KEIMARTELKLVANPRAVQEST 122  
QY 101 -----LATNTLDINSVLGKDYGALKDVIYINANPASPLSLVLRHLICERYVSLVH 153  
DB 123 PLKSKRAQLKAEPIINT-----CLQDFVNVASPEYVPSILALKNLWKRLNLQVECF 174  
QY 154 THSSVKNVPE-----NLVKCRGEOARKSRHEYDGLTLLWKVNPXTQMKFSVQTCPI 206  
DB 175 THSTVPKLSSEALAFQNAVNTASGTAAANLPR-----IKVLLIKVNGAVYEMTTSPTSYVP 230  
QY 207 IEGENIARFLFSL-----FGQKHNAVTLTLIDSWVDIANFOLREGSKKKAVERFMN 260  
DB 231 ICGEVNILRLYLSRCGPSEPNYEQDN---VDEVDLILDACVYLLKKNVKNQOQLRTLG 287  
QY 261 SALGSRPMLVGNELTVADVIMVSLQ 287  
DB 288 AKLGKRAAGFGADLSLCLDIATFSAVKQ 314  
  
RESULT 12  
Q7KUM5 PRELIMINARY; PRT; 301 AA.  
AC Q7KUM5;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE CG12304-PB.  
GN ORFName=CG12304;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gockyne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Suton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazet R.G., Chame M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abiri J.F., Agdayani A., An H.J., Andrews-Frankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,  
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Fábros B., Delcher A., Deng Z., Mays A.D., Dew I., Dierz S.M.,  
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison U.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Krat C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Meriklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Sampson M., Skupski M.P., Smith T.,  
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkask R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Ye J.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S.,  
RA Gibbs R.A., Myers E.W., Zhang G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RT *Science* 287:2185-2195 (2000).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537566;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lawley T., Muray D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svrtkask R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*  
RT *melanogaster* euchromatic genome sequence,"  
RL *Genome Biol.* 3:RESEARCH0079-RESEARCH0079 (2002).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svrtkask R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [4]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby W.A., Mungall C.U., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.,  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review,"  
RL *Genome Biol.* 3:RESEARCH0083-RESEARCH0083 (2002).  
RN [5]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426071; PubMed=12537574;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [6]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426072; PubMed=12537575;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [7]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426073; PubMed=12537576;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [8]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426074; PubMed=12537577;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [9]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426075; PubMed=12537578;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [10]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426076; PubMed=12537579;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [11]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426077; PubMed=12537580;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [12]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426078; PubMed=12537581;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [13]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426079; PubMed=12537582;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [14]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426080; PubMed=12537583;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [15]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426081; PubMed=12537584;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [16]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426082; PubMed=12537585;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [17]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426083; PubMed=12537586;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [18]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426084; PubMed=12537587;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [19]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426085; PubMed=12537588;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [20]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426086; PubMed=12537589;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [21]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426087; PubMed=12537590;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [22]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426088; PubMed=12537591;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [23]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426089; PubMed=12537592;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [24]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426090; PubMed=12537593;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [25]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426091; PubMed=12537594;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [26]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426092; PubMed=12537595;  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomic perspective,"  
RL *Genome Biol.* 3:RESEARCH0084-RESEARCH0084 (2002).  
RN [27]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=22426093; PubMed=12537596;  
RA Ashburner M., Celniker S.E.,  
RT "The trans

Qy	3	MYOVKPYHGSGAPLRVLEPTCMYRLPNV-----HSKTTSPATDAGHVOETSPSLQA	54
Db	1	MYELKTL---LPQDILPTCMYRLKNVSLAADSLSGSSTSASISA---STSCDDTAS	54
Qy	55	LESRQDDILKRLVELKAAVDGLSKMHTHPDADLVNTILOADEPTLATNTLDLSVLGK	114
Db	55	VVAHQEVLKQLELTKAQGLQIR-----AGLGVCG--KTFQHTTAFQNG-----GL	98
Qy	115	DYGLAKDIVINANASPLSLVYLRLLCGRYRLSLVHTHSSVKNV-----PENLVK	167
Db	99	KEVPLQDVVINGHNFPFLPYALDLAKNMRNLVYTDVKTFTHTSVADIGPAAFEANLAK	156
Qy	168	CFEGHQAKSGSRHEHVGFTLIMKNVPTQMKFVSQVTCMPCIEEGGNIAFLFSLFGOKHNA	227
Db	159	VPVNPALP-----KISVTLIMKNCERTHEMISPTWVPIYGEFNIIIRYLGRVGPARYR	212
Qy	228	VITLTL---IDSWDIAMFQURESSSEKKAVPFSSANSALGRSPMLVGNELTVADVILMSV	284
Db	213	EGSPLNEIDLVDIDICYQLRCNTHKTQVAMVRLDRLOKQOYFGGSGMSVADVGYSS	277
Qy	285	L 285	
Db	273	L 273	

RESULT 13

ID	Q8T060	PRELIMINARY;	PRT;	313 AA.
AC	Q8T060;			
DT	01-JUN-2002 (T-EMBLrel. 21, Created)			
DT	01-MAR-2004 (T-EMBLrel. 26, Last sequence update)			
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)			
DE	LD2572P (Fragment).			
OS	ORFName=CG12304;			
GN	Drosophila melanogaster (Fruit Fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_Taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkely;			
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Drenth D., Farfan D., Frise E.,			
RA	George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,			
RA	Miranda A., Mungall C.J., Nunoo J., Paclob J., Pargas V., Park S.,			
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.B., Rudin G.M.,			
RA	Celniker S.,			
RL	Submitted (DEC-2003) to the EMBL/genbank/DBJ databases.			
DR	EMBL; AY069537; AAL39682.2; -			
DR	PIRbase; P8gn0036515; CG12304.			
DR	InterPro; IPR004046; GST_Cterm.			
DR	InterPro; IPR010987; GST_C_like.			
DR	Pfam; PF00043; GST_C_1.			
FT	NON_TER			
SEQ	SEQUENCE	313 AA;	34522 MW;	5AF9C801877F8531 CRC64;

Query Match 13.5%; Score 224; DB 2; Length 313;  
 Best Local Similarity 28.6%; Pred. No. 4.3e-10;  
 Matches 86; Conservative 47; Mismatches 122; Indels 46; Gaps

```
QY 168 CFGEQARKSRHEVOLGFTLWKNVPTKTMKFSVOTMCEIEGEGNIARELFSLFGQKHNA 227
DB 171 VPVNRLP-----KISVTLWKNCHENTMISSPTMYVVIYGEVNIIRILGCVGPAEYRY 224
QY 228 VTLTL---IDSVNDIAMFOLREGSSKEKAVFRSNNSALGRSPWLGVNELTVADVVLMSV 284
DB 225 BSSPLCNEIDLVDICYQLLRGNTHTKTQVAMVRLDLKRLKQOQYFGCSQMSVADYGVSS 284
QY 285 L 285
DB 285 L 285

RESULT 14
Q6NKM4 PRELIMINARY; PRT; 322 AA.
ID Q6NKM4
AC Q6NKM4;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE LP12114P (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celinker S.;
RA Submitted (May-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BT012670; AAT08476.1; -
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C-like.
DR Pfam; PF00043; GST_C_1.
FT NON_TER
SQ SEQUENCE 322 AA; 35528 MW; 82A15927DF928D8AE CRC64;

Query Match 13.5%; Score 224; DB 2; Length 322;
Best Local Similarity 28.6%; Pred. No. 4,5e-10;
Matches 86; Conservative 47; Mismatches 122; Indels 46; Gaps 9;

QY 3 MYQVPEYHGSAPLVELEPTCMYRLPNV-----HSTTSPADNAGHVQTSSESLQA 54
DB 22 MYELKTL---LPQFDIKLPTCMYPLKNVSLADSLASGSSTASISA---STSSCDDTAS 75
QY 55 LESRODDILKRLYLEKAAVDSKMTHTPDADLDVTNIIQADBPPTLATNTDLNSLVGK 114
DB 76 VAAREBKVLKQLELKQAGQIR-----AGLGVG--KTRQHTTAPONG-----GL 119
QY 115 DYGALKDIVINANPASPRLSLVLRHLLCERYRVLTSTHTSSVKNV-----PENLVK 167
DB 120 KEVPLDQVIVNGHPNPIFYALLALKNAMRNLYTIDVKTFTHTMADIGPARREFEANLAK 179
QY 168 CEGEQARKSRHEVOLGFTLWKNVPTKTMKFSVOTMCEIEGEGNIARELFSLFGQKHNA 227
DB 180 VPVNRLP-----KISVTLWKNCHENTMISSPTMYVVIYGEVNIIRILGCVGPAEYRY 233
QY 228 VTLTL---IDSVNDIAMFOLREGSSKEKAVFRSNNSALGRSPWLGVNELTVADVVLMSV 284
DB 224 BSSPLCNEIDLVDICYQLLRGNTHTKTQVAMVRLDLKRLKQOQYFGCSQMSVADYGVSS 293
QY 285 L 285
DB 294 L 294

RESULT 15
MC22_DROME STANDARD; PRT; 334 AA.
ID MC22_DROME
AC Q9VUR3;
DT 16-OCT-2001 (Rel. 40, Created)
```

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Probable multisynthetase complex auxiliary component p38.
DE ORFNames=CG12304;
GN ORFNames=CG12304;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Galibert W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switkes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN [2]
RP GENE REANNOTATION.
RX MEDLINE=22426069; PubMed=12537572;
RA Miura S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RN Genome Biol. 3:RSEARCH0083.1-RSEARCH0083.22(2002).
CC -!- FUNCTION: Probable core protein of the multisynthetase complex
CC that serves as a template for the assembly of the supramolecular
CC structure (By similarity).
CC -!- SUBUNIT: Component of the multisynthetase complex which is
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the
CC monospecific isoleucyl, leucyl, glutamyl, methionyl, Iysyl,
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary
CC proteins, p18, p48 and p43 (By similarity).
CC -!- SIMILARITY: Contains 1 GST-like domain.
```

```

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ch/announce/
CC      or send an email to license@ebi.ebi.ch).
CC
CC      EMBL; AB003530; AAF49612.1; -.
CC      IntAct; Q9VUR3; -.
CC      FLYbase; FBgn0036515; CG12304.
CC      InterPro; IPR010987; GST_C1like.
CC      InterPro; IPR004046; GST_C1term.
CC      Pfam; PF00043; GST_C; 1.
CC      Protein biosynthesis.
CC      KW DOMAIN 280 327 GST-1like.
CC      SEQUENCE 334 AA; 36933 MW; B68FD70AE621990F CRC64;
SQ
Query Match      13.1%; Score 216.5; DB 1; Length 334;
Best Local Similarity 26.0%; Pred. No. 2e-09;
Matches 86; Conservative 48; Mismatches 124; Indels 73; Gaps 9;
QY      3 MYQVKPYHGGSAFLRVELPTCMRLPNV-----HSKTTSPATDAG----- 42
DB      1 MYELKTL--LPOFDIKLPTCMYPLKNVSLAADSLASGSSTASSTSSCKLEANRID 57
QY      43 -----HVQETSEPSLQALSESRODDILKRLVELKRAVDGLSKMHTPD 84
DB      58 RTGRNATCATLDDSLGRQIQRLKODPTASVAAKQERVLKQLELKAQLGQIR----- 110
QY      85 ADDLDVTNIIQADEPTTLATNTLDLNSVLGKQYALKDVIYINANPASPLSLVLRHLC 144
DB      111 AGIGVCG--KTFQHTTAFQNG-----GLKEVPLQDVINGHPNFIYALLALKNAMRN 161
QY      145 RYRVVSTVHTHSSVKV-----PENLYKCFGEQARKOSRHEVQLGFTLIMKNVPTQM 197
DB      162 LYTIIDVKTFTHTSMADIGPAAREFEANLAKVPVNPALP-----KISVTLIMKNCETEM 215
QY      198 KFSVQTMCPLEGEGNIARFLFSLFGQKHNAVTLTL--IDSWYDIAMFQLRGSSSEKAA 254
DB      216 ISSPTMYVPIYGEVNIIRYLGRVGAPEYRYESGSPLCNEIDVLDICYQLRCNTHKTQYA 275
QY      255 VFRSMNSALGRSPVLVGNELTVADVVLMSVL 285
DB      276 MVRLLDKRLQKQYFEGSQMSVADVGVYSSL 306

```

Search completed: February 23, 2005, 13:57:15  
 Job time : 61.9811 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:25:39 ; Search time 102.258 Seconds  
(without alignment)  
1758.725 Million cell updates/sec

Title: US-10-622-817-8  
Perfect score: 2596  
Sequence: 1 MIVFVRFNSSHGFPVEVDS...CNCGCEMNRVCMGDHWDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq.16Dec04:\*  
1: geneseqp19808:\*  
2: geneseqp19908:\*  
3: geneseqp20008:\*  
4: geneseqp20018:\*  
5: geneseqp20028:\*  
6: geneseqp20038:\*  
7: geneseqp20038:\*  
8: geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2596	100.0	465	2	AAV32501 Human par
2	2596	100.0	465	6	AAE30800 Human par
3	2586	100.0	465	6	ABO07156 Human p53
4	2584	99.5	465	6	AAE30801 Human par
5	2405	92.6	437	2	AAV32502 Human par
6	2405	92.6	437	6	ABO07157 Human p53
7	2170.5	83.6	464	4	AAE67517 Amino aci
8	2165.5	83.4	464	4	AAE67531 Amino aci
9	2165.5	83.4	464	4	AAE67532 Amino aci
10	2087.5	80.4	451	4	AAE67533 Amino aci
11	1696.5	65.4	316	6	ABO07158 Human p53
12	1679.5	64.7	344	4	AAE67521 Amino aci
13	1234.5	47.6	296	4	AAE67526 Amino aci
14	1058.5	40.8	468	4	ABE5114 Drosophil
15	1002	38.6	250	4	AAE67519 Amino aci
16	997	38.4	262	4	AAE67518 Amino aci
17	870	33.5	156	4	AAE67213 Amino aci
18	671.5	25.9	183	4	AAE67525 Amino aci
19	671.5	25.9	194	4	AAE67524 Amino aci
20	544.5	21.0	386	4	AAE6951 C. elegan
21	506	19.5	153	4	AAE67523 Amino aci
22	335	12.9	77	4	AAE67529 Amino aci
23	263	10.1	46	4	AAE17055 Peptide #
24	263	10.1	46	4	ABE36052 Peptide #
25	263	10.1	46	4	AAE29548 Peptide #

26	263	10.1	46	4	ABE30873 Peptide #
27	263	10.1	46	4	ABE21452 Protein #
28	263	10.1	46	4	AAE69222 Human bon
29	263	10.1	46	4	AAE56839 Human bra
30	263	10.1	46	4	ABE50898 Human liv
31	263	10.1	46	5	ABE38830 Human pep
32	252	9.7	106	8	ADN96489 Modified
33	251	9.7	63	4	AAE67522 Amino aci
34	251	9.7	105	4	AAE67520 Amino aci
35	251	9.7	105	8	ADN96487 Modified
36	240.5	9.3	503	4	ABE61708 Drosophil
37	240.5	9.3	503	7	ADN67082 Atrialde-1
38	240	9.2	511	4	ABE63665 Drosophil
39	239	9.2	520	8	ADP22562 Sea-squid
40	236.5	9.1	445	4	AAE21034 Human nuc
41	236.5	9.1	445	8	ADN96154 T cell ac
42	236.5	9.1	445	8	ADN96026 T cell ac
43	236.5	9.1	557	3	AAE98059 Human Rin
44	236.5	9.1	557	8	ADN96028 T cell ac
45	236	9.1	53	4	ABE67528 Amino aci

ALIGNMENTS

RESULT 1  
AAV32501  
ID AAV32501 standard; protein; 465 AA.  
AC AAV32501;  
XX  
DT 21-OCT-1999 (first entry)  
XX  
DE Human parkin gene variant protein.  
XX  
KW Parkinson's disease related gene; parkin gene; variant; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO940191-A1.  
XX  
PD 12-AUG-1999.  
XX  
PF 09-FEB-1999; 99WO-JP000545.  
XX  
PR 09-FEB-1998; 98JP-00027531.  
XX  
PA (SHIM/) SHIMIZU N.  
PA (MIZU/) MIZUNO Y.  
PI Shimizu N, Mizuno Y;  
XX  
DR WPI; 1999-494295/41.  
DR N-PSDB; AAX99923.  
XX  
PT Gene implicated in the pathology of Parkinson's disease, used for  
PT treatment of the disease.  
PS  
PS Claim 1; Page 83-88; 114pp; English.  
XX  
XX This sequence is encoded by a gene of the invention, and is implicated in  
XX the pathology of Parkinson's disease. This sequence is a variant of the  
XX parkin gene found in parkinson's disease patients. The sequences may be  
XX used for the diagnosis, treatment (including gene therapy) and  
XX investigation of Parkinson's disease  
SQ Sequence 465 AA;  
Query Match 100.0%; Score 2596; DB 2; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.3e-219; Indels 0; Gaps 0;  
Matches 465; Conservative 0; Mismatches 0;  
QY 1 MIVFVRFNSSHGFPVEVDSITFQKEVVAKRGVPADQLRVIFAGKELRNDWTQVNCND 60

```
Db 1 MIVFVFNSSHGPFVVDSDTSIFQLKEVVAKRQGVPAQDLVITAGEKELRNDWTYQNC 60
Qy 61 LDQOSIVHIYVQPMWKGEMNATGDDPRNAAGCEREPQSLTRVDLSSVLPEDSVGLA 120
Db 61 LDQOSIVHIYVQPMWKGEMNATGDDPRNAAGCEREPQSLTRVDLSSVLPEDSVGLA 120
Qy 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVRQPGKLRVQCSCTCROATLTLTQGP 180
Db 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVRQPGKLRVQCSCTCROATLTLTQGP 180
Qy 181 SCMDVLLIPNRMSEGCQSPHCPGTSAPFFKCGAHPSTDKETPVALLIATNSRNTICIT 240
Db 181 SCMDVLLIPNRMSEGCQSPHCPGTSAPFFKCGAHPSTDKETPVALLIATNSRNTICIT 240
Qy 241 CTDVRSPLYVFOQNSRHVYCLDPCFHLVCTRLNDQFVHDQGLSLPCVAGCNSLIKE 300
Db 241 CTDVRSPLYVFOQNSRHVYCLDPCFHLVCTRLNDQFVHDQGLSLPCVAGCNSLIKE 300
Qy 301 LHHFRILGEBQYNYRQYGAEECVLQMGVLCPRPGCAGLLPBPQKRYTCGGNGLGC 360
Db 301 LHHFRILGEBQYNYRQYGAEECVLQMGVLCPRPGCAGLLPBPQKRYTCGGNGLGC 360
Qy 361 GFAPCRECKEAYHGECSAVFEASGTTTQAYRVDERAABQARWEAASKETIKKTTKPCR 420
Db 361 GFAPCRECKEAYHGECSAVFEASGTTTQAYRVDERAABQARWEAASKETIKKTTKPCR 420
Qy 421 CHVPEKNGGCMHMKCPQPCRLEMCNCGCEWNRVCMGDHWFV 465
Db 421 CHVPEKNGGCMHMKCPQPCRLEMCNCGCEWNRVCMGDHWFV 465
```

## RESULT 2

```
AAB30800
ID AAB30800 standard; protein; 465 AA.
AC AAB30800;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human Parkin protein.
XX
KW Human; Parkin protein; neurological disorder; apoptosis; gene therapy;
KW ischaemic stroke; Parkinson's disease; Alzheimer's disease; noctropic;
KW transgenic; cerebroprotective; neuroprotective; neurotransplantation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 126
PN MO200279459-A2.
XX
PD 10-OCT-2002.
XX
PF 02-APR-2002; 2002WO-DK000221.
XX
PR 29-MAR-2001; 2001DK-00000525.
PR 03-APR-2001; 2001US-0281286P.
XX
PA (NSGE-) NSGENE AS.
XX
PI Jensen PH;
XX
DR WPI; 2003-046812/04.
DR N-PSDB; AAD47679.
XX
PT New isolated nucleic acid sequence encoding a Parkin polypeptide, useful
PT for treating, preventing or diagnosing neurological disorders, e.g.
PT Parkinson's disease, Alzheimer's disease or ischemic stroke, and in
PT screening assays.
XX
PS Claim 10; Page 69; 71pp; English.
```

```
XX
CC The invention relates to Parkin protein and its corresponding nucleic
CC acid sequence. The nucleic acid sequence is useful for altering the
CC proteolytic processing of Parkin at its potential cleavage site at Asp
CC 126. The invention is used in manufacturing or testing a pharmaceutical
CC composition for treating and/or preventing a neurological disorder, e.g.
CC Alzheimer's disease or ischaemic stroke. It also used for detecting the
CC occurrence of proteolytic processing of Parkin at Asp 126 in a sample, in
CC monitoring a potential disposition for a neurodegenerative disease and
CC for treating, preventing and/or diagnosing Parkinson's disease or other
CC neurodegenerative disorders. The viral vector is used for transforming
CC neuronal cells in vivo or ex vivo. The invention is useful for
CC neurotransplantation into the CNS of a mammal. It may be used in
CC screening assays to identify compounds that increase or decrease
CC apoptosis. It is also used in gene therapy. The present sequence is human
CC Parkin protein
XX
SQ Sequence 465 AA;
```

Query Match 100.0%; Score 2596; DB 6; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.3e-219;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MIVFVFNSSHGPFVVDSDTSIFQLKEVVAKRQGVPAQDLVITAGEKELRNDWTYQNC 60
Db 1 MIVFVFNSSHGPFVVDSDTSIFQLKEVVAKRQGVPAQDLVITAGEKELRNDWTYQNC 60
Qy 61 LDQOSIVHIYVQPMWKGEMNATGDDPRNAAGCEREPQSLTRVDLSSVLPEDSVGLA 120
Db 61 LDQOSIVHIYVQPMWKGEMNATGDDPRNAAGCEREPQSLTRVDLSSVLPEDSVGLA 120
Qy 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVRQPGKLRVQCSCTCROATLTLTQGP 180
Db 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVRQPGKLRVQCSCTCROATLTLTQGP 180
Qy 181 SCMDVLLIPNRMSEGCQSPHCPGTSAPFFKCGAHPSTDKETPVALLIATNSRNTICIT 240
Db 181 SCMDVLLIPNRMSEGCQSPHCPGTSAPFFKCGAHPSTDKETPVALLIATNSRNTICIT 240
Qy 241 CTDVRSPLYVFOQNSRHVYCLDPCFHLVCTRLNDQFVHDQGLSLPCVAGCNSLIKE 300
Db 241 CTDVRSPLYVFOQNSRHVYCLDPCFHLVCTRLNDQFVHDQGLSLPCVAGCNSLIKE 300
Qy 301 LHHFRILGEBQYNYRQYGAEECVLQMGVLCPRPGCAGLLPBPQKRYTCGGNGLGC 360
Db 301 LHHFRILGEBQYNYRQYGAEECVLQMGVLCPRPGCAGLLPBPQKRYTCGGNGLGC 360
Qy 361 GFAPCRECKEAYHGECSAVFEASGTTTQAYRVDERAABQARWEAASKETIKKTTKPCR 420
Db 361 GFAPCRECKEAYHGECSAVFEASGTTTQAYRVDERAABQARWEAASKETIKKTTKPCR 420
Qy 421 CHVPEKNGGCMHMKCPQPCRLEMCNCGCEWNRVCMGDHWFV 465
Db 421 CHVPEKNGGCMHMKCPQPCRLEMCNCGCEWNRVCMGDHWFV 465
```

## RESULT 3

```
AAB007156
ID AAB007156 standard; protein; 465 AA.
AC AAB007156;
XX
DT 13-AUG-2003 (first entry)
XX
DE Human p53 modifying protein, SEQ ID 116.
XX
KW Human; p53 modifier; cyrostatic; cancer; cytostatic; antiangiogenic;
KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
KW apoptotic disorder; cell proliferation disorder.
XX
OS Homo sapiens.
XX
```

XX	PN	MO200299122-AA.
XX	PD	
XX	FD	12-DEC-2002.
XX	PF	03-JUN-2002; 2002WO-US017382.
XX	PR	05-JUN-2001; 2001US-0296076P.
XX	PR	10-OCT-2001; 2001US-0328605P.
XX	PR	15-FEB-2002; 2002US-0357253P.
XX	PA	(EXEL-) EXELIXIS INC.
XX	PI	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP,
XX	DR	WPI; 2003-156859/15.
XX	DR	N-PDSB; ACDI3332.
PT	PT	Identifying modulators of the p53 pathway for use in treating apoptotic
PT	PT	or cell proliferation disorders, comprises screening for agents that
PT	PT	modulate activity of a human ortholog of genes that modify the p53
PT	PT	pathway in Drosophila.
PS	PS	Example 2; Page 399-401; 678bp; English.
CC	CC	The invention relates to identifying (M1) a candidate p53 pathway
CC	CC	modulating agent, by contacting an assay system comprising a purified HM
CC	CC	polypeptide (human orthologue of genes that modify the p53 pathway in
CC	CC	Drosophila) or nucleic acid with a test agent under conditions, where but
CC	CC	for the presence of the test agent, the system provides a reference
CC	CC	activity, and detecting a test agent-biased activity of the assay system.
CC	CC	Also included are modulating (M2) a p53 pathway of a cell (comprising
CC	CC	contacting a cell defective in p53 function with a candidate modulator
CC	CC	that specifically binds to a HM polypeptide comprising an HM amino acid
CC	CC	sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC	CC	in a mammalian cell (comprising contacting the cell with an agent that
CC	CC	specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC	CC	a disease in a patient (comprising: (a) obtaining a biological sample
CC	CC	from the patient); (b) contacting the sample with a probe for HM
CC	CC	expression; (c) comparing the results with a control; and (d) determining
CC	CC	whether the comparison indicates a likelihood disease). (M1) is useful
CC	CC	for identifying modulators of the p53 pathway. A probe for HM expression
CC	CC	is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC	CC	in a patient, where the cancer has greater than 25 % expression level.
CC	CC	Modulators identified by (M1) are useful in a variety of diagnostic and
CC	CC	therapeutic applications, where disease or disorder prognosis is related
CC	CC	to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC	CC	proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC	CC	M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC	CC	the p53 function of the cell, so that the cell undergoes normal
CC	CC	proliferation or progression through the cell cycle. (M2) and (M3) are
CC	CC	also useful for treating defects in the p53 pathway such as angiogenic,
CC	CC	apoptotic or cell proliferation disorders. The present sequence
CC	CC	represents a human p53 pathway modifying protein
XX	XX	
SQ	SQ	Sequence 465 AA;
		Query Match            100.0%; Score 2596; DB 6; Length 465;
		Best Local Similarity   100.0%; Pred. No. 2.3e-219;
		Matches   465; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
OY	OY	1 MIVFRENSSHGFPVEVDSDTISIFOLKEVNAKRQGVPADQLRVIPAGKELRNDWTVGNCDD
DB	DB	1 MIVFVRFNSSHGFPEVVDSDTISIFOLKEVNAKRQGVPADQLRVIPAGKELRNDWTVGNCDD
OY	OY	61 LDGGSIVHIIVQRPAKGGDMNATGDDDDPRNAAAGCGEREPQSITRDVLSSVLPDGSVGLA 120
DB	DB	61 LDGGSIVHIIVQRPMKGGEMNATGDDDDPRNAAAGCGEREPQSITRDVLSSSVLPDGSVGLA 120
OY	OY	121 VILHTDSRKSDSPAPSGASRIYSFYCYCKGPCQCVOPGKLRYVCSTCRQAATLTLTGGP 180
DB	DB	121 VILHTDSRKSDSPAPSGASRIYSFYCYCKGPCQCVOPGKLRYVCSTCRQAATLTLTGGP 180
OY	OY	181 SCWDVLLINRMSEGCQSPHCPTSABEFFKCAHPSTDKETPVALLIATNSRNITCIT 240

Db	181	SCMDVLIIPRMSEGCSPHCPEGTSAEFFPKCANPISDKETVALHLIATNSRNTTCIT	240
Qy	241	CTDVRSPVLVFOCNSRHVICLDGFHLVCTVRLNDROFVHDPOLGYSLPVAGCPNSLIXE	300
Db	241	CTDVRSPVLVFOCNSRHVICLDGFHLVCTVRLNDROGVHDPOLGYSLPVAGCPNSLIXE	300
Qy	301	LHHRRIIGEEQYNNRYQYGAEECYLQMGVYLCPRPCCGAGLPEPDRKVTCEGNGLGC	360
Db	301	LHHRRIIGEEQYNNRYQYGAEECYLQMGVYLCPRPCCGAGLPEPDRKVTCEGNGLGC	360
Qy	361	GFACRCKEAKYHGECSAIVEASGTTQAYRDERBAEQARWEASKETIKKTKPCPR	420
Db	361	GFACRCKEAKYHGECSAIVEASGTTQAYRDERBAEQARWEASKETIKKTKPCPR	420
Qy	421	CHVVEKNKGCMHMKCPQPCRLMECMNCGEENRYCMGDHMDV	465
Db	421	CHVVEKNKGCMHMKCPQPCRLMECMNCGEENRYCMGDHMDV	465
RESULT 4			
AAE30801	ID	AAE30801 standard; protein; 465 AA.	
AC	AAE30801;		
XX			
DT	24-FEB-2003	(first entry)	
XX			
DE	Human Parkin D126 mutant protein.		
XX			
KW	Human; Parkin protein; neurological disorder; apoptosis; gene therapy;		
KW	ischemic stroke; Parkinson's disease; Alzheimer's disease; neurotic;		
KW	transgenic; cerebroprotective; neuroprotective; neurotransplantation;		
KW	mutant; mutein.		
XX			
OS	Homo sapiens.		
CS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 126		
FT	/note= "Wild-type Asp is replaced with Glu"		
FT	Cleavage-site 126		
FT	Misc-difference 223		
FT	/note= "Encoded by CCA"		
XX			
PN	WO200279459-A2.		
XX			
PD	10-OCT-2002.		
XX			
PF	02-APR-2002; 2002MO-DK000221.		
XX			
PR	29-MAR-2001; 2001DK-00000525.		
PR	03-APR-2001; 2001US-0281286P.		
XX			
PA	(NSGE-) NSGENE AS.		
XX			
PI	Jensen PH;		
XX			
DR	WPI; 2003-046812/04.		
DR	N-PSDB; AAD47680.		
XX			
PT	New isolated nucleic acid sequence encoding a Parkin polypeptide, useful for		
PT	treating, preventing or diagnosing neurological disorders, e.g.		
PT	Parkinson's disease, Alzheimer's disease or ischemic stroke, and in		
XX	screening assays.		
XX			
ES	Claim 10; Page 71; 71pp; English.		
XX			
CC	The invention relates to Parkin protein and its corresponding nucleic		
CC	acid sequence. The nucleic acid sequence is useful for altering the		
CC	proteolytic processing of Parkin at its potential cleavage site at Asp		
CC	126. The invention is used in manufacturing or testing a pharmaceutical		
CC	composition for treating and/or preventing a neurological disorder, e.g.		





XX WO200239122-A1.  
 PN  
 XX 12-DEC-2002.  
 PD  
 XX 03-JUN-2002; 2002MO-US017382.  
 PF  
 XX 05-JUN-2001; 2001US-0296076P.  
 PR 10-OCT-2001; 2001US-028605P.  
 PR 15-FEB-2002; 2002US-0357253P.  
 XX (EXEL-) EXELIXIS INC.  
 PA  
 PI Friedman L, Plowman GD, Belvin W, Francis-Lang H, Li D, Funke RP;  
 XX  
 XX MPI: 2003-156859/15.  
 DR N-PSDB; ACD13333.  
 XX  
 PT Identifying modulators of the p53 pathway for use in treating apoptotic  
 PT or cell proliferation disorders, comprises screening for agents that  
 PT modulate activity of a human ortholog of genes that modify the p53  
 PT pathway in Drosophila.

Example 2; Page 401-402; 678bp; English.

CC The invention relates to identifying (M1) a candidate p53 pathway  
 CC modulating agent, by contacting an assay system comprising a purified HM  
 CC polypeptide (human orthologue of genes that modify the p53 pathway in  
 CC Drosophila) or nucleic acid with a test agent under conditions, where but  
 CC for the presence of the test agent, the system provides a reference  
 CC activity, and detecting a test agent-biased activity of the assay system.  
 CC Also included are modulating (M2) a p53 pathway of a cell (comprising  
 CC contacting a cell defective in p53 function with a candidate modulator  
 CC that specifically binds to a HM polypeptide comprising an HM amino acid  
 CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
 CC in a mammalian cell (comprising contacting the cell with an agent that  
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
 CC a disease in a patient (comprising: (a) obtaining a biological sample  
 CC from the patient; (b) contacting the sample with a probe for HM  
 CC expression; (c) comparing the results with a control; and (d) determining  
 CC whether the comparison indicates a likelihood disease). (M1) is useful  
 CC for identifying modulators of the p53 pathway. A probe for HM expression  
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
 CC in a patient, where the cancer has greater than 25 % expression level.  
 CC Modulators identified by (M1) are useful in a variety of diagnostic and  
 CC therapeutic applications, where disease or disorder prognosis is related  
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
 CC the p53 function of the cell, so that the cell undergoes normal  
 CC proliferation or progression through the cell cycle. (M2) and (M3) are  
 CC also useful for treating defects in the p53 pathway such as angiogenic,  
 CC apoptotic or cell proliferation disorders. The present sequence  
 CC represents a human p53 pathway modifying protein

Sequence 437 AA;

Query Match 92.6%; Score 2405; DB 6; Length 437;

Best Local Similarity 94.0%; Pred. No. 1.3e-202; Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MIVPFRPNSSHGPPVVDSDTSIFOLKEVAVKQGVADOLRIYFAGKELRNMTWQND 60  
 DB 1 MIVPFRPNSSHGPPVVDSDTSIFOLKEVAVKQGVADOLRIYFAGKELRNMTWQND 60  
 QY 61 LDOQSIIVHIVORPWKQCEMNATGDDPRNAAGCGEREPOSTLTVLSSVLPJGDSVGLA 120  
 DB 61 LDOQSIIVHIVORPWKQCEMNATGDDPRNAAGCGEREPOSTLTVLSSVLPJGDSVGLA 120  
 QY 121 VILHTDSRKDSPPAGSPAGRSIYNSFYVCKGFCQGVQVQPKLRVQCSTGRQATLTITGCP 180  
 DB 121 VILHTDSRKDSPPAGSPAGRSIYNSFYVCKGFCQGVQVQPKLRVQCSTGRQATLTITGCP 180  
 QY 121 VILHTDSRKDSPPAGSPAGRSIYNSFYVCKGFCQGVQVQPKLRVQCSTGRQATLTITGCP 180  
 DB 121 VILHTDSRKDSPPAGSPAGRSIYNSFYVCKGFCQGVQVQPKLRVQCSTGRQATLTITGCP 180

QY 181 SCMDVLLIPNRMSGECQSPHCPTSAEFFKCGAKPTSKETPVAILIATNSRNTICIT 240  
 DB 179 -----EFFKCGAKPTSKETPVAILIATNSRNTICIT 212  
 QY 241 CTDVASPVTVFOCNSRHVCLDCPFHLXCYTRLRNDQFVNDPQGVSLPCVAGCPNLIIE 300  
 DB 213 CTDVASPVTVFOCNSRHVCLDCPFHLXCYTRLRNDQFVNDPQGVSLPCVAGCPNLIIE 272  
 QY 301 LHHFRILGEOYNNRYOQYGAEECVLQMGVLCPRPCGAGLLPEPQRYVTEGGNGLGC 360  
 DB 273 LHHFRILGEOYNNRYOQYGAEECVLQMGVLCPRPCGAGLLPEPQRYVTEGGNGLGC 332  
 QY 361 GFAPCECKEAYHEGCSAVFEASGTTTQAYVDERBAEQARWEAASKETIKKTYPCPR 420  
 DB 333 GFAPCECKEAYHEGCSAVFEASGTTTQAYVDERBAEQARWEAASKETIKKTYPCPR 392  
 QY 421 CHVPEYKNGGCMHMKCPQPCRLKEMWCMNCGCEMNRVCMGDHMFV 465  
 DB 393 CHVPEYKNGGCMHMKCPQPCRLKEMWCMNCGCEMNRVCMGDHMFV 437

RESULT 7

AAB67517  
 ID AAB67517 standard; protein; 464 AA.

AC AAB67517;  
 XX

DT 29-MAY-2001 (first entry)  
 XX

DE Amino acid sequence of a murine parkin2 polypeptide.  
 XX

KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW Brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.  
 XX

OS Mus sp.  
 XX

PN EP1081225-A1.  
 XX

PD 07-MAR-2001.  
 XX

PF 30-AUG-1999; 99EP-00116766.  
 XX

PR 30-AUG-1999; 99EP-00116766.  
 XX

PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX

PI Luebbert H;  
 XX

DR MPI, 2001-212797/22.  
 XX

DR N-PSDB; AAF55244.  
 XX

PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 XX

PS Disclosure; Page 17-19; 62pp; English.  
 XX

XX The present sequence represents a murine parkin2 polypeptide. Mutations  
 CC or deletions in the parkin2 gene cause Parkinson's disease in humans. The  
 CC human parkin2 gene is located in gene region 6q25.2-27. Parkin2  
 CC polypeptides and polynucleotides are useful for analysing  
 CC neurodegenerative diseases. They are also useful for testing the efficacy  
 CC of the treatment of a neurodegenerative disease such as Parkinson's  
 CC disease, Alzheimer's disease, Huntington's disease, amyotrophic lateral  
 CC sclerosis, Multi-system atrophy, Wilson's disease, Pick's disease, Prion  
 CC disease, and secondary causes inducing Parkinson's syndromes like toxins,  
 CC drugs, brain tumours, head trauma, stroke, vascular irregularities or  
 CC metabolic irregularities, associated with a less active or non-active  
 CC parkin protein  
 XX

SQ Sequence 464 AA;  
 Query Match 83.6%; Score 2170.5; DB 4; Length 464;  
 Best Local Similarity 83.4%; Pred. No. 66-182;  
 Matches 388; Conservative 29; Mismatches 47; Indels 1; Gaps 1;

QY 1 MIVFVFNSSHGPFPEVDSSTSI FOLKEVNAKRGVADQLRVIPAGKELRNDMTVQNC D 60  
 DB 1 MIVFVFNSSHGPFPEVDSSTSI FOLKEVNAKRGVADQLRVIPAGKELRNDMTVQNC D 60  
 QY 61 LDOGSIVHIVORPMKCGEMNATGDDPRNAGCEREPOSITRVDSSTVLPDSDVGLA 120  
 DB 61 LDOGSIVHIVORPMKCGEMNATGDDPRNAGCEREPOSITRVDSSTVLPDSDVGLA 120  
 QY 121 VILHTDSKRDSPAPGSPAGRSIYNSFYVYCKGPCQRYVQGLRVQSCSTCRQATLTLLQGP 180  
 DB 121 VILHTDSKRDSPAPGSPAGRSIYNSFYVYCKGPCQRYVQGLRVQSCSTCRQATLTLLQGP 180  
 QY 121 VILHTDSKRDSEARGPV-KPTVNSFFIYCKGPCCHKYQGLRVQSCSTCRQATLTLLQGP 179  
 DB 121 VILHTDSKRDSEARGPV-KPTVNSFFIYCKGPCCHKYQGLRVQSCSTCRQATLTLLQGP 179  
 QY 181 SCMDVLLIPNRMSSGCGSPHCPGTSAEFFPKCGAHPISDKETPVALLIATNSRNTCTT 240  
 DB 181 SCMDVLLIPNRMSSGCGSPHCPGTSAEFFPKCGAHPISDKETPVALLIATNSRNTCTT 240  
 QY 241 CTDVRSPLVLFQCNRRHVICLDCEFLHYCVTRLNDROFVHDPOGLGYSLPCVAGCPSNLIXE 300  
 DB 241 CTDVRSPLVLFQCNRRHVICLDCEFLHYCVTRLNDROFVHDPOGLGYSLPCVAGCPSNLIXE 300  
 QY 301 LHHFRILGEEQYRNYOYGAEECVLQMGVLCPRPGCGAGLLPEPDQRKVTCEGGNGLGC 360  
 DB 301 LHHFRILGEEQYRNYOYGAEECVLQMGVLCPRPGCGAGLLPEPDQRKVTCEGGNGLGC 360  
 QY 361 GFACRECKEAYHGECSAFAEASGTTQAYRVDERAAEQARWEASKETIKKTTTPCPR 420  
 DB 361 GFACRECKEAYHGECSAFAEASGTTQAYRVDERAAEQARWEASKETIKKTTTPCPR 420  
 QY 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 465  
 DB 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 465  
 QY 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 464  
 DB 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 464

RESULT 8  
 AAB67531 ID AAB67531 standard; protein; 464 AA.  
 AC AAB67531;  
 XX  
 DT 29-MAY-2001 (first entry)  
 DE Amino acid sequence of a mutated murine parkin2 polypeptide.  
 XX  
 KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.  
 XX  
 OS Mus sp.  
 XX  
 PN EP1081225-A1.  
 XX  
 PD 07-MAR-2001.  
 XX  
 PF 30-AUG-1999; 99EP-00116766.  
 XX  
 PR 30-AUG-1999; 99EP-00116766.  
 XX  
 PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX  
 PI Luebbert H;  
 DE  
 XX  
 XX WPI; 2001-212797/22.  
 DR N-PSDB; AAF55258.  
 XX  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing

PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PS  
 XX Claim 7; Page 47-49; 62pp; English.  
 CC The present sequence represents a murine parkin2 polypeptide. The  
 CC sequence contains the mutation Lys161Asn. Mutations or deletions in the  
 CC parkin2 gene cause Parkinson's disease in humans. The human parkin2 gene  
 CC is located in gene region 6q25.2-27. Parkin2 polypeptides and  
 CC polynucleotides are useful for analysing neurodegenerative diseases. They  
 CC are also useful for testing the efficacy of the treatment of a  
 CC neurodegenerative disease such as Parkinson's disease, Alzheimer's  
 CC disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-  
 CC system atrophy, Wilson's disease, Pick's disease, Prion disease, and  
 CC secondary causes inducing Parkinson's syndrome like toxins, drugs, brain  
 CC tumours, head trauma, stroke, vascular irregularities or metabolic  
 CC irregularities, associated with a less active or non-active parkin  
 CC protein  
 CC  
 XX  
 SQ Sequence 464 AA;  
 Query Match 83.4%; Score 2165.5; DB 4; Length 464;  
 Best Local Similarity 83.2%; Pred. No. 1.7e-181;  
 Matches 387; Conservative 29; Mismatches 48; Indels 1; Gaps 1;

QY 1 MIVFVFNSSHGPFPEVDSSTSI FOLKEVNAKRGVADQLRVIPAGKELRNDMTVQNC D 60  
 DB 1 MIVFVFNSSHGPFPEVDSSTSI FOLKEVNAKRGVADQLRVIPAGKELRNDMTVQNC D 60  
 QY 61 LDOGSIVHIVORPMKCGEMNATGDDPRNAGCEREPOSITRVDSSTVLPDSDVGLA 120  
 DB 61 LDOGSIVHIVORPMKCGEMNATGDDPRNAGCEREPOSITRVDSSTVLPDSDVGLA 120  
 QY 121 VILHTDSKRDSPAPGSPAGRSIYNSFYVYCKGPCQRYVQGLRVQSCSTCRQATLTLLQGP 180  
 DB 121 VILHTDSKRDSPAPGSPAGRSIYNSFYVYCKGPCQRYVQGLRVQSCSTCRQATLTLLQGP 180  
 QY 121 VILHTDSKRDSEARGPV-KPTVNSFFIYCKGPCCHKYQGLRVQSCSTCRQATLTLLQGP 179  
 DB 121 VILHTDSKRDSEARGPV-KPTVNSFFIYCKGPCCHKYQGLRVQSCSTCRQATLTLLQGP 179  
 QY 181 SCMDVLLIPNRMSSGCGSPHCPGTSAEFFPKCGAHPISDKETPVALLIATNSRNTCTT 240  
 DB 181 SCMDVLLIPNRMSSGCGSPHCPGTSAEFFPKCGAHPISDKETPVALLIATNSRNTCTT 240  
 QY 241 CTDVRSPLVLFQCNRRHVICLDCEFLHYCVTRLNDROFVHDPOGLGYSLPCVAGCPSNLIXE 300  
 DB 241 CTDVRSPLVLFQCNRRHVICLDCEFLHYCVTRLNDROFVHDPOGLGYSLPCVAGCPSNLIXE 300  
 QY 301 LHHFRILGEEQYRNYOYGAEECVLQMGVLCPRPGCGAGLLPEPDQRKVTCEGGNGLGC 360  
 DB 301 LHHFRILGEEQYRNYOYGAEECVLQMGVLCPRPGCGAGLLPEPDQRKVTCEGGNGLGC 360  
 QY 361 GFACRECKEAYHGECSAFAEASGTTQAYRVDERAAEQARWEASKETIKKTTTPCPR 420  
 DB 361 GFACRECKEAYHGECSAFAEASGTTQAYRVDERAAEQARWEASKETIKKTTTPCPR 420  
 QY 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 465  
 DB 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 465  
 QY 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 464  
 DB 421 CHVPEKNGGCMAMKCPQPCRLKEMCWNCGCEMNRVCMGDHMFV 464

RESULT 9  
 AAB67532 ID AAB67532 standard; protein; 464 AA.  
 AC AAB67532;  
 XX  
 DT 29-MAY-2001 (first entry)  
 DE Amino acid sequence of a mutated murine parkin2 polypeptide.  
 XX  
 KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.

XX	Mus sp.
OS	EPI081225-A1.
XX	
PN	07-MAR-2001.
XX	
PD	30-AUG-1999; 99BP-00116766.
XX	
PB	30-AUG-1999; 99BP-00116766.
XX	
PR	(BIOF-) BIOFRONTIERA PHARM GMBH.
PA	Luebbert H;
PL	WPI; 2001-212797/22.
XX	N-PDSB; AAF55259.
DR	New polynucleotides encoding mouse parkin protein, useful for producing
PT	a transgenic non-human animal as an animal model for neurodegenerative
FT	diseases.
PT	
PS	Claim 7; Page 49-51; 62pp; English.
XX	
CC	The present sequence represents a murine parkin2 polypeptide. The
CC	sequence contains the mutation Thr415Asn. Mutations or deletions in the
CC	parkin gene cause Parkinson's disease in humans. The human parkin2 gene
CC	is located in gene region 6q25.2-27. Parkin2 polypeptides and
CC	polynucleotides are useful for analysing neurodegenerative diseases. They
CC	are also useful for testing the efficacy of the treatment of a
CC	neurodegenerative disease such as Parkinson's disease, Alzheimer's
CC	disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-
CC	system atrophy, Wilson's disease, Pick's disease, prion disease, and
CC	secondary causes inducing Parkinson's syndromes like toxins, drugs, brain
CC	tumours, head trauma, stroke, vascular irregularities or metabolic
CC	irregularities, associated with a less active or non-active parkin
CC	protein
XX	
SQ	Sequence 464 AA;
	Query Match      83.4%; Score 2165.5; DB 4; Length 464;
	Best Local Similarity    83.2%, Pred. No. 17e-181; Mismatches 48; Indels 1; Gaps 1
	Matches 387; Conservative 29; Mismatched 48; Indels 1; Gaps 1
OY	1 MIVFREFFSSHGFPVEVDSDTISIFLKEVAVAKGVAPADQLRVIPAGKELRNDWTIQRNCD 60
Dd	1 MIVFRFRFSSTGFPVEVDSDTSIIDLKEVAAKGVPADQLRVIRAGKELRNHLVLQNCD 60
OY	61 LDQSIIHVORPWRKKGMATAGDDPPRNAAGCEREPOSITRVDISSSVLPDGSVGIA 120
Dd	61 LEOGSIIHVORPRRRSHETNASGDDEPOSTSEGISIWESRLTRVDLSHTLTPVDSVGIA 120
OY	121 VILHRDKSDSPPASPGASRTSYNSFYCYCKCPQVRQGKLRYOCCTCRQTTLTLTGGR 180
Dd	121 VILDTRSRSDBEAAGPV-KPTVNSFFTYCKGPCHKVOPGKLRYOCCGTCKQTTLTLAQGP 179
OY	181 SCMDVLLPNNMSGECQSPHCPTSAEFPFKCAHPSTSOKETPEVALMLTIATSNITCT 240
Dd	180 SCMDVLLPNNMSGECQSPDCGTAEFPFKCAHPSTDXTSVALNLITSNRKRIPCIA 239
OY	241 CTDVARSPLYVFQCNSRHVICLDPCFLYCVTRLNDRQVFNHDPLGYSLPCVACSPNSLIKE 300
Dd	240 CTDVARSPLYVFQCNSRHVICLDPCFLYCVTRLNDRQVFNHDGLYSGLPCVACSPNSLIKE 299
OY	301 LHHPRIILEGEQYNRYOYGAEFCVLOMGCVLCPRGCCAGLLPEDQRKYTCGGNGIGIC 360
Dd	300 LHHPRIILEGEQYTRYOYGAEFCVLOMGCVLCPRGCCAGLLPEGQQRKYTCGGNGIGIC 359
OY	361 GFAPRECEKAYHEEGSCSAVFASGTTTOAYVDERAAEQAMEAASKETIKTKTCPER 420
Dd	360 GFVPERDCKEVAHEEDCSLSLEPSATSQAYVYDDRABAEQAWEBAASKETIKTKTCPER 419
OY	421 CHIVEKNKGCGMNMKCPOQPQCRLEMCMNCGCSEMNVFCMGDMFDV 465

[illegible]

Db 121 VILDTSRKDSSEARGPV-KPTVNSFFIYCKGPHKVGKLRVQCGTKCATLTLAGP 179  
 QY 181 SCMDVDLIPRMSSGEGSPHCPTSAEFPFKGAPHTSPKTPVALHLIATNSRNTTCT 240  
 Db 180 SCMDVDLIPRMSSGEGSPHCPTSAEFPFKGAPHTSPKTPVALHLIATNSRNTTCT 240  
 QY 241 CTDVRSPLVIFQCNRSRNVICLDGPHLYCTRLNDROFVHDPOGLGSLPCVAGCPNSLKE 300  
 Db 240 CTDVRSPLVIFQCNRSRNVICLDGPHLYCTRLNDROFVHDPOGLGSLPCVAGCPNSLKE 299  
 QY 301 LHHFRILGEEQYRNYOQYGAECVLOMGVLCPRPGCAGLILPEPDQRKVTCEGNGIGC 360  
 Db 300 LHHFRILGEEQYRNYOQYGAECVLOMGVLCPRPGCAGLILPEPDQRKVTCEGNGIGC 359  
 QY 361 GFAPCRCKEAYHGECSAFVFEASGTTTQAYRVDERAAEOARWEASKEITIKTTKPCPR 420  
 Db 360 GFAPCRCKEAYHGECSAFVFEASGTTTQAYRVDERAAEOARWEASKEITIKTTKPCPR 419  
 QY 421 CHVPEKNGGCMHMKCPQPCRLKEMWNCGE 452  
 Db 420 CNVPEKNGGCMHMKCPQPCRLKEMWNCGE 451

RESULT 11

ABO07158 ID ABO07158 standard; protein; 316 AA.

XX AC ABO07158;

XX DT 13-AUG-2003 (first entry)

DE Human p53 modifying protein, SEQ ID 118.

XX KM Human; p53 modifier; cytoskeletal; cancer; cytoskeletal; antiangiogenic;  
 KM antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
 KM lung cancer; ovarian cancer; angiogenesis; cell cycle;  
 KM apoptotic disorder; cell proliferation disorder.

OS Homo sapiens.

XX PN W0200299122-A1.

XX PD 12-DEC-2002.

XX PF 03-JUN-2002; 2002W0-US017382.

XX PR 05-JUN-2001; 2001US-0296076P.

XX PR 10-OCT-2001; 2001US-0328605P.

XX PR 15-FEB-2002; 2002US-0357253P.

XX PA (EXEL-) EXELIXIS INC.

XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX DR WPI; 2003-156859/15.

XX DR N-PSDB; ACD13334.

XX PT Identifying modulators of the p53 pathway for use in treating apoptotic

XX PT or cell proliferation disorders, comprises screening for agents that

XX PT modulate activity of a human ortholog of genes that modify the p53

XX PT pathway in Drosophila.

XX PS Example 2; Page 402-403; 678bp; English.

XX CC The invention relates to identifying (M1) a candidate p53 pathway  
 CC modulating agent, by contacting an assay system comprising a purified HM  
 CC polypeptide (human ortholog of genes that modify the p53 pathway in  
 CC Drosophila) or nucleic acid with a test agent under conditions, where but  
 CC for the presence of the test agent, the system provides a reference  
 CC activity, and detecting a test agent-biased activity of the assay system.  
 CC Also included are modulating (M2) a p53 pathway of a cell comprising  
 CC contacting a cell defective in p53 function with a candidate modulator  
 CC that specifically binds to a HM polypeptide comprising an HM amino acid

CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
 CC in a mammalian cell (comprising contacting the cell with an agent that  
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
 CC a disease in a patient (comprising: (a) obtaining a biological sample  
 CC from the patient; (b) contacting the sample with a probe for HM  
 CC expression; (c) comparing the results with a control; and (d) determining  
 CC whether the comparison indicates a likelihood of disease). (M1) is useful  
 CC for identifying modulators of the p53 pathway. A probe for HM expression  
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer.  
 CC In a patient, where the cancer has greater than 25 % expression level.  
 CC Modulators identified by (M1) are useful in a variety of diagnostic and  
 CC therapeutic applications, where disease or disorder prognosis is related  
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
 CC the p53 function of the cell, so that the cell undergoes normal  
 CC proliferation or progression through the cell cycle. (M2) and (M3) are  
 CC also useful for treating defects in the p53 pathway such as angiogenic,  
 CC apoptotic or cell proliferation disorders. The present sequence  
 CC represents a human p53 pathway modifying protein

XX S0 Sequence 316 AA;

Query Match 65.4%; Score 1696.5; DB 6; Length 316;

Best Local Similarity 68.0%; Pred. No. 1.8e-140; Mismatches 0; Indels 149; Gaps 1;

Matches 316; Conservative 0; Mismatches 0; Indels 149; Gaps 1;

QY 1 MIVFVRENSHGFPVVDSDTSIFOLKEVYAKRGVADQLRVIFAGKELENDTVONCD 60

Db 1 MIVFVRENSHGFPVVDSDTSIFOLKEVYAKRGVADQLRVIFAGKELENDTVONCD 57

QY 61 LDQGSIVHIVPRKQENMATGDDPRNAGCEREPQSLRTVDLSSVLPDSVGLA 120

Db 58 ----- 57

QY 121 VILHTRKDSPPAGSPAGRSIYNSFYVCKGPGQVRVQPKLRVQCGSTCRQATLTLLQGP 180

Db 58 ----- 57

QY 181 SCMDVDLIPRMSSGEGSPHCPTSAEFPFKGAPHTSPKTPVALHLIATNSRNTTCT 240

Db 58 -----EFPFKGAPHTSPKTPVALHLIATNSRNTTCT 91

QY 241 CTDVRSPLVIFQCNRSRNVICLDGPHLYCTRLNDROFVHDPOGLGSLPCVAGCPNSLKE 300

Db 92 CTDVRSPLVIFQCNRSRNVICLDGPHLYCTRLNDROFVHDPOGLGSLPCVAGCPNSLKE 151

QY 301 LHHFRILGEEQYRNYOQYGAECVLOMGVLCPRPGCAGLILPEPDQRKVTCEGNGIGC 360

Db 152 LHHFRILGEEQYRNYOQYGAECVLOMGVLCPRPGCAGLILPEPDQRKVTCEGNGIGC 211

QY 361 GFAPCRCKEAYHGECSAFVFEASGTTTQAYRVDERAAEOARWEASKEITIKTTKPCPR 420

Db 212 GFAPCRCKEAYHGECSAFVFEASGTTTQAYRVDERAAEOARWEASKEITIKTTKPCPR 271

QY 421 CHVPEKNGGCMHMKCPQPCRLKEMWNCGE 465

Db 272 CHVPEKNGGCMHMKCPQPCRLKEMWNCGE 316

RESULT 12

AB67521 ID AAB67521 standard; protein; 344 AA.

XX AC AAB67521;

XX DT 29-MAY-2001 (first entry)

XX DE Amino acid sequence of a murine truncated parkin2 polypeptide.

XX KM Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KM Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KM Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;

KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.  
 OS Mus sp.  
 XX EPI081225-A1.  
 XX PD 07-MAR-2001.  
 XX PF 30-AUG-1999; 99EP-00116766.  
 XX PR 30-AUG-1999; 99EP-00116766.  
 XX PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX PI Luebert H;  
 XX DR WPI; 2001-212797/22.  
 XX DR N-PSDB; AAF55248.  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PS Claim 7; Page 40-41; 62pp; English.  
 CC The present sequence represents a murine parkin2 polypeptide. The  
 CC polynucleotide sequence contains a deletion, leading to a truncated  
 CC protein. Mutations or deletions in the parkin2 gene cause Parkinson's  
 CC disease in humans. The human parkin2 gene is located in gene region  
 CC 6q25.2-27. Parkinson polypeptides and polynucleotides are useful for  
 CC analysing neurodegenerative diseases. They are also useful for testing  
 CC the efficacy of the treatment of a neurodegenerative disease such as  
 CC Parkinson's disease, Alzheimer's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,  
 CC Pick's disease, Prion disease, and secondary causes inducing Parkinson's  
 CC syndromes like toxins, drugs, brain tumours, head trauma, stroke,  
 CC vascular irregularities or metabolic irregularities, associated with a  
 CC less active or non-active parkin protein  
 XX Sequence 344 AA;  
 SQ  
 Query Match 64.7%; Score 1679.5; DB 4; Length 344;  
 Best Local Similarity 66.0%; Pred. No. 6.3e-139;  
 Matches 307; Conservative 14; Mismatches 23; Indels 121; Gaps 1;

QY 1 MIVFVRENSHGFPVEVSDTSIFOLKEVAVARQGVADQLRVIFAGKELRNDWTYQNC 60  
 DB 1 MIVFVRENSHGFPVEVSDTSIFOLKEVAVARQGVADQLRVIFAGKELRNDWTYQNC 60  
 QY 61 LDOQSTVHIVQRPWRKGGEMNATGGDDPRNNAAGCEREPQSLTRVLDSSVLRPGDSVGLA 120  
 DB 57 ----- 56  
 QY 121 VILHTDSRKDSPAGSPAGRSIYNFVYCKGPCQHVQPKLKVQSTCRQATLTLTQGP 180  
 DB 57 ----- 56  
 QY 57 ----- 59  
 DB 181 SCWDDVLIENRMSGEGSCSPHCPGTSAPFFKCGAMPTSDKETPVALLIATNSRNTTCT 240  
 DB 60 SCWDDVLIENRMSGEGSCSPHCPGTSAPFFKCGAMPTSDKETPVALLIATNSRNTTCT 240  
 QY 241 CTDVNSPVVFOCNSRHVCLDCCFHLXYCTRLNDRPFVNDPQUGYSIPCVAGCPNLLIKE 300  
 DB 120 CTDVNSPVVFOCNSRHVCLDCCFHLXYCTRLNDRPFVNDPQUGYSIPCVAGCPNLLIKE 300  
 QY 301 LHHFRILGEEQVNRVYQOYGAEBECVLMQGVLCPRPGCAGLLPEPDQRKYTCGGNGLGC 360  
 DB 180 LHHFRILGEEQVNRVYQOYGAEBECVLMQGVLCPRPGCAGLLPEPDQRKYTCGGNGLGC 360  
 QY 361 GAFCPECKEAYVHEGSCAVFEASGTTTQAYRVDERBAQAQWEAASKETIKKTKPCPR 420  
 DB 240 GAFCPECKEAYVHEGSCAVFEASGTTTQAYRVDERBAQAQWEAASKETIKKTKPCPR 420

QY 421 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEWNRVCMGDHMDV 465  
 DB 300 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEWNRVCMGDHMDV 344  
 RESULT 13  
 ID AAB67526 standard; protein; 296 AA.  
 AC AAB67526;  
 DT 29-MAY-2001 (first entry)  
 XX Amino acid sequence of a murine truncated parkin2 polypeptide.  
 KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.  
 XX Mus sp.  
 OS EPI081225-A1.  
 XX PD 07-MAR-2001.  
 XX PF 30-AUG-1999; 99EP-00116766.  
 XX PR 30-AUG-1999; 99EP-00116766.  
 XX PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX PI Luebert H;  
 XX DR WPI; 2001-212797/22.  
 XX DR N-PSDB; AAF55253.  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PS Claim 7; Page 44-45; 62pp; English.  
 CC The present sequence represents a murine parkin2 polypeptide. The  
 CC polynucleotide sequence contains a deletion, leading to a truncated  
 CC protein. Mutations or deletions in the parkin2 gene cause Parkinson's  
 CC disease in humans. The human parkin2 gene is located in gene region  
 CC 6q25.2-27. Parkinson polypeptides and polynucleotides are useful for  
 CC analysing neurodegenerative diseases. They are also useful for testing  
 CC the efficacy of the treatment of a neurodegenerative disease such as  
 CC Parkinson's disease, Alzheimer's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,  
 CC Pick's disease, Prion disease, and secondary causes inducing Parkinson's  
 CC syndromes like toxins, drugs, brain tumours, head trauma, stroke,  
 CC vascular irregularities or metabolic irregularities, associated with a  
 CC less active or non-active parkin protein  
 XX Sequence 296 AA;  
 SQ  
 Query Match 47.6%; Score 1234.5; DB 4; Length 296;  
 Best Local Similarity 80.0%; Pred. No. 7.3e-100;  
 Matches 232; Conservative 20; Mismatches 37; Indels 1; Gaps 1;

QY 1 MIVFVRENSHGFPVEVSDTSIFOLKEVAVARQGVADQLRVIFAGKELRNDWTYQNC 60  
 DB 1 MIVFVRENSHGFPVEVSDTSIFOLKEVAVARQGVADQLRVIFAGKELRNDWTYQNC 60  
 QY 61 LDOQSTVHIVQRPWRKGGEMNATGGDDPRNNAAGCEREPQSLTRVLDSSVLRPGDSVGLA 120  
 DB 61 LDOQSTVHIVQRPWRKGGEMNATGGDDPRNNAAGCEREPQSLTRVLDSSVLRPGDSVGLA 120



CC sclerosis, Multi-system atrophy, Wilson's disease, Prion  
CC disease, and secondary causes inducing Parkinson's syndromes like toxins,  
CC drugs, brain tumours, head trauma, stroke, vascular irregularities or  
CC metabolic irregularities, associated with a less active or non-active  
CC parkin protein

XX Sequence 250 AA;

Query Match 38.6%; Score 1002; DB 4; Length 250;

Best Local Similarity 77.6%; Pred. No. 1.7e-79;

Matches 190; Conservative 20; Mismatches 35; Indels 0; Gaps 0;

Qy	1	MIVFVRFNSSHGPFVEVDSDTSIFQLKEVAVAKQGVPAQQLRVIPAGKILRNDMTVQNC	60
		:     :     :     :     :     :     :     :     :	
Db	1	MIVFVRFNSSYGPFVEVDSDTSILQLKEVAVAKQGVPAQQLRVIPAGKILPNHLTVQNC	60
		:     :     :     :     :     :     :     :	
Qy	61	LDQOSIVHIVQRPWRKQGMNATGDDPRNNAAGCCEREPOSILTRYDLSSTVLPDPSVGLA	120
		:     :     :     :     :     :     :     :     :	
Db	61	LEQOSIVHIVQRPWRSHETNASGDEPOSTEGSIWESRSLTRYDLSSTVLPDPSVGLA	120
		:     :     :     :     :     :     :     :     :	
Qy	121	VILHTRSRKDSPPAGSPAGRSIYNSFYVYCKGPCQRVOPGKLRVOCSTCROATLTLTQGP	180
		:     :     :     :     :     :     :     :	
Db	121	VILDTSRKDSFAAGPAVKPTYNPFITYCKGPCHKVQPGKLRVOCGTCKQATLTLAGSP	180
		:     :     :     :     :     :     :     :	
Qy	181	SCWMDVLIIPNRMSGECOSPHCPGTSAEFFKCAHPTSDKETPVALHLIATNSRNITCIT	240
		:     :     :     :     :     :     :     :	
Db	181	SCWMDVLIIPNRMSGECQSPDCGTRAEFFKCAHPTSDKTSVALNLITSNRRSIPCTA	240
		:     :     :     :     :     :     :     :	
Qy	241	CTDVR 245	
Db	241	CTDVR 245	

Search completed: February 23, 2005, 13:50:12  
Job time : 105.258 secs

iris page blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:36:20 ; Search time 25.8801 Seconds  
(without alignments)  
1341.256 Million cell updates/sec

Title: US-10-622-817-8

Perfect score: 2596  
Sequence: 1 MIVFVRFNSSHGFPVEVDS.....CNCGCCENRVCQGHMFDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfillseq.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596	100.0	465	US-09-601-844B-2	Sequence 2, Appl1
2	2596	100.0	465	US-09-949-016-6516	Sequence 6516, Ap
3	2405	92.6	437	US-09-601-844B-4	Sequence 4, Appl1
4	2405	92.6	437	US-09-949-016-6853	Sequence 6853, Ap
5	1696.5	65.4	316	US-09-949-016-6854	Sequence 6854, Ap
6	370.5	14.3	117	US-09-270-767-32183	Sequence 32183, A
7	240.5	9.3	503	US-09-914-259-13	Sequence 13, Appl1
8	222	8.6	474	US-09-354-221-2	Sequence 2, Appl1
9	222	8.6	487	US-09-949-016-10111	Sequence 10111, A
10	191	7.4	506	US-09-248-796A-15410	Sequence 15410, A
11	167	6.4	587	US-08-398-008A-2	Sequence 2, Appl1
12	167	6.4	587	US-08-893-333-2	Sequence 2, Appl1
13	149	5.7	464	US-09-538-092-598	Sequence 598, App
14	141	5.4	328	US-09-248-796A-16471	Sequence 16471, A
15	133	5.1	352	US-08-854-764-2	Sequence 2, Appl1
16	133	5.1	352	PCT-US95-09377-2	Sequence 2, Appl1
17	131.5	5.1	1106	US-09-949-016-9626	Sequence 9626, Ap
18	130.5	5.0	1242	US-09-488-270A-2	Sequence 2, Appl1
19	128.5	4.9	229	US-08-726-306A-23	Sequence 23, Appl1
20	128.5	4.9	229	US-08-840-126-20	Sequence 20, Appl1
21	128.5	4.9	229	US-09-360-220-20	Sequence 20, Appl1
22	128	4.9	160	US-09-370-838-205	Sequence 205, App
23	128	4.9	160	US-09-854-133-205	Sequence 205, App
24	127	4.9	2321	US-09-230-652-2	Sequence 2, Appl1
25	126	4.9	237	US-09-248-796A-19062	Sequence 19062, A
26	126	4.9	846	US-08-149-103-3	Sequence 3, Appl1
27	126	4.9	846	US-08-451-883-3	Sequence 3, Appl1

28	126	4.9	873	1	US-08-393-734-2	Sequence 2, Appl1
29	126	4.9	973	3	US-08-894-489-2	Sequence 2, Appl1
30	126	4.9	904	4	US-09-949-016-9528	Sequence 9528, Ap
31	125.5	4.8	103	2	US-08-771-201-9	Sequence 9, Appl1
32	125	4.8	76	1	US-08-232-815-2	Sequence 2, Appl1
33	125	4.8	76	1	US-08-350-906-2	Sequence 2, Appl1
34	125	4.8	76	1	US-09-601-844B-61	Sequence 61, Appl1
35	125	4.8	76	5	PCT-US95-04536-2	Sequence 2, Appl1
36	125	4.8	206	4	US-09-248-796A-19430	Sequence 19430, A
37	125	4.8	1121	1	US-07-789-915A-2	Sequence 2, Appl1
38	125	4.8	1121	1	US-08-005-002C-2	Sequence 2, Appl1
39	125	4.8	1121	1	US-08-487-203A-2	Sequence 2, Appl1
40	124.5	4.8	77	6	5510474-3	Patent No. 5510474
41	124.5	4.8	77	6	5510474-3	Patent No. 5510474
42	124.5	4.8	305	4	US-09-866-153-6	Sequence 6, Appl1
43	124.5	4.8	305	4	US-09-693-467A-6	Sequence 6, Appl1
44	124.5	4.8	305	4	US-09-270-976-6	Sequence 6, Appl1
45	124.5	4.8	381	4	US-09-866-153-9	Sequence 9, Appl1

## ALIGNMENTS

RESULT 1									
US-09-601-844B-2									
Sequence 2, Application US/09601844B									
Patent No. 6716621									
GENERAL INFORMATION:									
APPLICANT: Shimizu, No. 6716621uyoshi									
APPLICANT: Mizuno, Yoshiaki									
TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease									
FILE REFERENCE: 0652.2110000									
CURRENT APPLICATION NUMBER: US/09/601,844B									
PRIOR FILING DATE: 2000-08-09									
PRIOR APPLICATION NUMBER: PCT/JP99/00545									
PRIOR FILING DATE: 1999-02-09									
NUMBER OF SEQ ID NOS: 70									
SOFTWARE: Patentin version 3.1									
SEQ ID NO 2									
LENGTH: 465									
TYPE: PRT									
ORGANISM: Homo sapiens									
US-09-601-844B-2									
Query Match									
Best Local Similarity 100.0%; Score 2596; DB 4; Length 465;									
Matches 465; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MIVFVRFNSSHGFPVEVDS	DTSI	1	FOLEKVVAKRGVPADQLRV	IFAGKEIRNDWVQNC	60		
DB	1	MIVFVRFNSSHGFPVEVDS	DTSI	1	FOLEKVVAKRGVPADQLRV	IFAGKEIRNDWVQNC	60		
QY	61	LDQOSTVHIVQRPWRKGQEMNATGDDPRNAA	GCEREPQSLTRVDLSSVLP	FGDSVGLA	120				
DB	61	LDQOSTVHIVQRPWRKGQEMNATGDDPRNAA	GCEREPQSLTRVDLSSVLP	FGDSVGLA	120				
QY	121	VILHDSRKDPSPPAGSRISYNSFYVCKP	CCQVQVQGXLRVQSTORQATLT	TOGP	180				
DB	121	VILHDSRKDPSPPAGSRISYNSFYVCKP	CCQVQVQGXLRVQSTORQATLT	TOGP	180				
QY	181	SCMDVLIENRNSGECQSPHCPTSAEPFK	CAHPTSKETPVVAHLATNSRNT	TCIT	240				
DB	181	SCMDVLIENRNSGECQSPHCPTSAEPFK	CAHPTSKETPVVAHLATNSRNT	TCIT	240				
QY	241	CTDVRSPLVLFQCNRSRHVILCD	CFHLVYCTRINDRQFVHD	POLGYSLPCVAGCPNS	LIKE	300			
DB	241	CTDVRSPLVLFQCNRSRHVILCD	CFHLVYCTRINDRQFVHD	POLGYSLPCVAGCPNS	LIKE	300			
QY	301	LHHFRLTGEQVNRQVGAEECVLQ	MPREGCGALLPEPQRYVTG	CGGNGLGC	360				
DB	301	LHHFRLTGEQVNRQVGAEECVLQ	MPREGCGALLPEPQRYVTG	CGGNGLGC	360				
QY	361	GFAFCRECKEAVHGECSAVFEAS	GTTTQAYRVDERRAQAQAEAS	SKETIKKTKTPCPR	420				
DB	361	GFAFCRECKEAVHGECSAVFEAS	GTTTQAYRVDERRAQAQAEAS	SKETIKKTKTPCPR	420				

Db 361 GFAFCECKEAYHEGCSAVFEASGTTTAYVDERAAEQARWEASKEITKTKTCPR 420  
QY 421 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 465  
Db 421 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 465

## RESULT 2

US-09-949-016-6516  
Sequence 6516, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949, 016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 2000-09-08  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 6516  
LENGTH: 465  
TYPE: PR  
ORGANISM: Human  
US-09-949-016-6516

Query Match 100.0%; Score 2596; DB 4; Length 465;  
Best Local Similarity 100.0%; Pred. No. 8.4e-251;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVADQLRVIFAGKEIRNDMTVONCD 60  
Db 1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVADQLRVIFAGKEIRNDMTVONCD 60  
QY 61 LDQOSIVHIVQPRMKGEMNATGDDPRNAAAGCCEREPOSILTRVDSLSSVLPQDSVGLA 120  
Db 61 LDQOSIVHIVQPRMKGEMNATGDDPRNAAAGCCEREPOSILTRVDSLSSVLPQDSVGLA 120  
QY 121 VILHDSKSDPPAGSPAGRSIYNSFYVCKGPGCORVQPGKLRVQSTCRQATLTITQGP 180  
Db 121 VILHDSKSDPPAGSPAGRSIYNSFYVCKGPGCORVQPGKLRVQSTCRQATLTITQGP 180  
QY 181 SCMDVLIIPNMSGECSPHCPGTSAEFFPKCGAHTSDKETPVALLIATNSRNITCIT 240  
Db 181 SCMDVLIIPNMSGECSPHCPGTSAEFFPKCGAHTSDKETPVALLIATNSRNITCIT 240  
QY 241 CTDRSPVLVFOQNSRHVILCLDFHLYCVTRLNDRQFVADPOLGYSILPCVAGCPNSLIKE 300  
Db 241 CTDRSPVLVFOQNSRHVILCLDFHLYCVTRLNDRQFVADPOLGYSILPCVAGCPNSLIKE 300  
QY 301 LHHFRILGEEQYRKYQOYGAEECVLQMGVLCPRPGCAGLLPEBDQKVTCEGNGIGC 360  
Db 301 LHHFRILGEEQYRKYQOYGAEECVLQMGVLCPRPGCAGLLPEBDQKVTCEGNGIGC 360  
QY 361 GFAFCECKEAYHEGCSAVFEASGTTTQAYVDERAAEQARWEASKEITKTKTCPR 420  
Db 361 GFAFCECKEAYHEGCSAVFEASGTTTQAYVDERAAEQARWEASKEITKTKTCPR 420  
QY 421 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 465  
Db 421 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 465

## RESULT 3

US-09-601-844B-4  
Sequence 4, Application US/09601844B

Patent No. 6716621  
GENERAL INFORMATION:  
APPLICANT: Shimizu, No. 6716621yueohi  
APPLICANT: Mizuno, Yoshiaki  
TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease  
FILE REFERENCE: 0652.211000  
CURRENT APPLICATION NUMBER: US/09/601,844B  
PRIOR FILING DATE: 2000-08-09  
PRIOR APPLICATION NUMBER: PCT/JP99/00545  
PRIOR FILING DATE: 1999-02-09  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 437  
TYPE: PR  
ORGANISM: Homo sapiens  
US-09-601-844B-4

Query Match 92.6%; Score 2405; DB 4; Length 437;  
Best Local Similarity 94.0%; Pred. No. 9.6e-232;  
Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVADQLRVIFAGKEIRNDMTVONCD 60  
Db 1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVADQLRVIFAGKEIRNDMTVONCD 60  
QY 61 LDQOSIVHIVQPRMKGEMNATGDDPRNAAAGCCEREPOSILTRVDSLSSVLPQDSVGLA 120  
Db 61 LDQOSIVHIVQPRMKGEMNATGDDPRNAAAGCCEREPOSILTRVDSLSSVLPQDSVGLA 120  
QY 121 VILHDSKSDPPAGSPAGRSIYNSFYVCKGPGCORVQPGKLRVQSTCRQATLTITQGP 180  
Db 121 VILHDSKSDPPAGSPAGRSIYNSFYVCKGPGCORVQPGKLRVQSTCRQATLTITQGP 180  
QY 181 SCMDVLIIPNMSGECSPHCPGTSAEFFPKCGAHTSDKETPVALLIATNSRNITCIT 240  
Db 179 -----EFPKCGAHTSDKETPVALLIATNSRNITCIT 212  
QY 241 CTDRSPVLVFOQNSRHVILCLDFHLYCVTRLNDRQFVADPOLGYSILPCVAGCPNSLIKE 300  
Db 213 CTDRSPVLVFOQNSRHVILCLDFHLYCVTRLNDRQFVADPOLGYSILPCVAGCPNSLIKE 272  
QY 301 LHHFRILGEEQYRKYQOYGAEECVLQMGVLCPRPGCAGLLPEBDQKVTCEGNGIGC 360  
Db 273 LHHFRILGEEQYRKYQOYGAEECVLQMGVLCPRPGCAGLLPEBDQKVTCEGNGIGC 332  
QY 361 GFAFCECKEAYHEGCSAVFEASGTTTQAYVDERAAEQARWEASKEITKTKTCPR 420  
Db 333 GFAFCECKEAYHEGCSAVFEASGTTTQAYVDERAAEQARWEASKEITKTKTCPR 392  
QY 421 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 465  
Db 393 CHVPEKNGGCMHMKCPQPCRLKEMCNGCEMNRVCMDHMFV 437

## RESULT 4

US-09-949-016-6853  
Sequence 6853, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949, 016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6853  
 ; LENGTH: 437  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-949-016-6853

Query Match 92.6%; Score 2405; DB 4; Length 437;  
 Best Local Similarity 94.0%; Pred. No. 9, 6e-232;  
 Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

```

QY 1 MIVPVRFNSSHGFPVEVSDTSIFOLKEVYAKQGVADQLRVIFAGKELRNDWTYQNC 60
DB 1 MIVPVRFNSSHGFPVEVSDTSIFOLKEVYAKQGVADQLRVIFAGKELRNDWTYQNC 60
QY 61 LDQOSIVHIVQRPKRGEMNATGGDDPRNNAAGCGEREPQSLTRVDLSSVLPBGDSVGLA 120
DB 61 LDQOSIVHIVQRPKRGEMNATGGDDPRNNAAGCGEREPQSLTRVDLSSVLPBGDSVGLA 120
QY 121 VILHTDSRKDSPAGSPAGRSIYNSFYVCKGPCQVRQPKLRVQSTCRQATLLTQGP 180
DB 121 VILHTDSRKDSPAGSPAGRSIYNSFYVCKGPCQVRQPKLRVQSTCRQATLLTQGP 180
QY 181 SCMDVLI.PNRMSGSCQSPHCPTSAEFPFKCGAHPDTSKETPVALHLIATNSRNITCIT 240
DB 181 SCMDVLI.PNRMSGSCQSPHCPTSAEFPFKCGAHPDTSKETPVALHLIATNSRNITCIT 240
QY 241 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 300
DB 241 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 300
QY 213 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 272
DB 213 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 272
QY 301 LHHFRILGEOYNYQOYGAEBCVLOMGVLCPRPCGAGLLPEPDRKYTCGEGNGLGC 360
DB 273 LHHFRILGEOYNYQOYGAEBCVLOMGVLCPRPCGAGLLPEPDRKYTCGEGNGLGC 332
QY 361 GFAPRCEKEAYHBEBCSAVFASGTTTQAYRVDERAABQAWBAASKETIKTTKPCR 420
DB 333 GFAPRCEKEAYHBEBCSAVFASGTTTQAYRVDERAABQAWBAASKETIKTTKPCR 392
QY 421 CHVPEKNGGCMHMKCPQPCRLWCWNCGCCMNRYCMGDHMFV 465
DB 393 CHVPEKNGGCMHMKCPQPCRLWCWNCGCCMNRYCMGDHMFV 437

```

RESULT 5  
 ; US-09-949-016-6854  
 ; Sequence 6854, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6854  
 ; LENGTH: 316  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-949-016-6854

Query Match 65.4%; Score 1696.5; DB 4; Length 316;  
 Best Local Similarity 68.0%; Pred. No. 4, 2e-161;  
 Matches 316; Conservative 0; Mismatches 0; Indels 149; Gaps 1;

```

QY 1 MIVPVRFNSSHGFPVEVSDTSIFOLKEVYAKQGVADQLRVIFAGKELRNDWTYQNC 60
DB 1 MIVPVRFNSSHGFPVEVSDTSIFOLKEVYAKQGVADQLRVIFAGKELRNDWTYQNC 57
QY 61 LDQOSIVHIVQRPKRGEMNATGGDDPRNNAAGCGEREPQSLTRVDLSSVLPBGDSVGLA 120
DB 58 ----- 57
QY 121 VILHTDSRKDSPAGSPAGRSIYNSFYVCKGPCQVRQPKLRVQSTCRQATLLTQGP 180
DB 58 ----- 57
QY 181 SCMDVLI.PNRMSGSCQSPHCPTSAEFPFKCGAHPDTSKETPVALHLIATNSRNITCIT 240
DB 58 ----- 57
QY 241 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 300
DB 92 CTDVSPVLVFOCNSRHVYCLDCFHLVCTYTRLNDQFVNDPOLGYSLPVACCPNSLIE 151
QY 301 LHHFRILGEOYNYQOYGAEBCVLOMGVLCPRPCGAGLLPEPDRKYTCGEGNGLGC 360
DB 152 LHHFRILGEOYNYQOYGAEBCVLOMGVLCPRPCGAGLLPEPDRKYTCGEGNGLGC 211
QY 361 GFAPRCEKEAYHBEBCSAVFASGTTTQAYRVDERAABQAWBAASKETIKTTKPCR 420
DB 212 GFAPRCEKEAYHBEBCSAVFASGTTTQAYRVDERAABQAWBAASKETIKTTKPCR 271
QY 421 CHVPEKNGGCMHMKCPQPCRLWCWNCGCCMNRYCMGDHMFV 465
DB 272 CHVPEKNGGCMHMKCPQPCRLWCWNCGCCMNRYCMGDHMFV 316

```

RESULT 6  
 ; US-09-270-767-32183  
 ; Sequence 32183, Application US/09270767  
 ; Patent No. 6703491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Homburger et al.  
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 ; FILE REFERENCE: File Reference: 7326-094  
 ; CURRENT APPLICATION NUMBER: US/09/270,767  
 ; CURRENT FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 62517  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 32183  
 ; LENGTH: 117  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-270-767-32183

Query Match 14.3%; Score 370.5; DB 4; Length 117;  
 Best Local Similarity 53.3%; Pred. No. 3, 8e-29;  
 Matches 64; Conservative 11; Mismatches 40; Indels 5; Gaps 2;

```

QY 345 PDQRKYTCGNGGLGFGAFRCEKEAYHBEBC-SAVFASGTTTQAYRVDERAABQAW 403
DB 1 PDQRKYTCGNGGLGFGAFRCEKEAYHBEBC-SAVFASGTTTQAYRVDERAABQAW 56
QY 404 EASAKETIKTTKPCRCHVPEKNGGCMHMKCPQPCRLWCWNCGCCMNRYCMGDHMF 463
DB 57 DEASNTVITKTPCKKCTPTERDGGCMHMKCTRAGCGFEMCWVQGTETMDCMGAWMF 116

```

RESULT 7  
 ; US-09-914-259-13  
 ; Sequence 13, Application US/09914259  
 ; Patent No. 6495336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makowski, Lee  
 ; APPLICANT: Hyman, Paul  
 ; APPLICANT: Williams, Mark  
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

FILE REFERENCE: 8471-010-999  
 CURRENT APPLICATION NUMBER: US/09/914,259  
 CURRENT FILING DATE: 2000-11-21  
 NUMBER OF SEQ ID NOS: 180  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 13  
 LENGTH: 503  
 TYPE: PRT  
 ORGANISM: Drosophila melanogaster  
 US-09-914-259-13

Query Match 9.3%; Score 240.5; DB 4; Length 503;  
 Best Local Similarity 27.8%; Pred. No. 3e-15;  
 Matches 72; Conservative 42; Mismatches 111; Indels 37; Gaps 13;

QY 209 PFKGAHPTSDKTEPVALHLIATNSRMITITCTDVRSPLVYQCNRSRVTCLDCEPHLYC 268  
 DB 105 FFKC-AHVNPFNAATEIKQKTRSCQCEBCEITFSQLPDMSMAGLECGHFCMPCMEYL 163  
 QY 269 VTRLANDQFVHDQLGSLPCLVA-GCPNSLIKELHFRILGEOYN-RYQYGAEECYLQ 326  
 DB 164 STK-----IVAGLQQTISCAHGC-DILVDVTVANLVTDARVAVYQQLITNSFY-E 215  
 QY 327 MGVLT-CPRPGCG-AGLPEPDQKVTGEGNGLCCGAFRCCKEAYHEG-ECSAVFE 382  
 DB 216 CNGLRMCPSTVDTYAVKPYAPRRVHCK-----CGHVFCAFGEMHDPVXCRWL- 267  
 QY 383 ASGTTQAVRVDRAAEQARWEASKEITIKTTKPCPRCHVPVEKNGGCMHKCPQPCR 442  
 DB 268 -----KKWTKCDDSETSNMIAAN-----TKSCPRCSVTIRKDGGMHVKCKNCK 315  
 QY 443 LEMCNCGEMNRVCMGDMHFD 464  
 DB 316 NEFCWVCLGSWEP-HGSSWYN 335

RESULT 8  
 US-09-354-221-2  
 Sequence 2, Application US/09354221  
 Patent No. 6699714  
 GENERAL INFORMATION:  
 APPLICANT: Chang, Chawmshang  
 TITLE OF INVENTION: Androgen Receptor Coactivators  
 FILE REFERENCE: 920920.90011  
 CURRENT APPLICATION NUMBER: US/09/354,221  
 CURRENT FILING DATE: 1999-07-15  
 EARLIER APPLICATION NUMBER: 60/100,243  
 EARLIER FILING DATE: 1998-09-14  
 NUMBER OF SEQ ID NOS: 12  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 2  
 LENGTH: 474  
 TYPE: PRT  
 ORGANISM: Homo sapien  
 US-09-354-221-2

Query Match 8.6%; Score 222; DB 4; Length 474;  
 Best Local Similarity 27.8%; Pred. No. 2e-13;  
 Matches 71; Conservative 27; Mismatches 97; Indels 60; Gaps 13;

QY 232 NSRNITC-ITCTDVSPVLVF-QCNSRHVCLDCEPHLYCVRRLNDROFVHDPLQGYSLP 288  
 DB 214 NSKLFCSICFCEKLSCECMYFLC-RHYCKACIKDYFEIQRDQ-----VQ 261  
 QY 289 CVAAGCN-----SLIKELHFRILGEOYNRYQYGAEECYLQMGVLT-CPRPGCG 338  
 DB 262 CL-NCEPRKPSVATGQVKEL-----VEALFARYDRLLQSSLDLMDADVTCPRPCCQ 315  
 QY 339 AGLPEPDQKVTGEGNGLCCGAFRCCKEAYH-EBCSAVFEASGTTQAY----- 391  
 DB 316 LPWQEPG-----CTWGLSSCNFAFCTLCLTYHGVSFCKVTAEKMLDRLNEYLQADEA 370  
 QY 392 -----RVDERAAEQARWEASKEITIKTTKPCPRCHVPVEKNGGCMHKCPQPCRLE 444

DB 371 NKRLLDQRYGRVYQKALEMESKEWLEKNSKSCPCGTPLEKLDGCMKMTG--TGCWQY 428  
 QY 445 WCMNCGEMNRVCMG 459  
 DB 429 FCW-----ICWG 435

RESULT 9  
 US-09-949-016-10111

Sequence 10111, Application US/09949016  
 Patent No. 6812339  
 GENERAL INFORMATION:  
 APPLICANT: VENTER, J. Craig et al.  
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 FILE REFERENCE: CL001307  
 CURRENT APPLICATION NUMBER: US/09/949,016  
 CURRENT FILING DATE: 2000-04-14  
 PRIOR APPLICATION NUMBER: 60/241,755  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/237,768  
 PRIOR FILING DATE: 2000-10-03  
 PRIOR APPLICATION NUMBER: 60/231,498  
 PRIOR FILING DATE: 2000-09-08  
 NUMBER OF SEQ ID NOS: 207012  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 10111  
 LENGTH: 487  
 TYPE: PRT  
 ORGANISM: Human  
 US-09-949-016-10111

Query Match 8.6%; Score 222; DB 4; Length 487;  
 Best Local Similarity 27.8%; Pred. No. 2e-13;  
 Matches 71; Conservative 27; Mismatches 97; Indels 60; Gaps 13;

QY 232 NSRNITC-ITCTDVSPVLVF-QCNSRHVCLDCEPHLYCVRRLNDROFVHDPLQGYSLP 288  
 DB 227 NSKLFCSICFCEKLSCECMYFLC-RHYCKACIKDYFEIQRDQ-----VQ 274  
 QY 289 CVAAGCN-----SLIKELHFRILGEOYNRYQYGAEECYLQMGVLT-CPRPGCG 338  
 DB 275 CL-NCEPRKPSVATGQVKEL-----VEALFARYDRLLQSSLDLMDADVTCPRPCCQ 328  
 QY 339 AGLPEPDQKVTGEGNGLCCGAFRCCKEAYH-EBCSAVFEASGTTQAY----- 391  
 DB 329 LPWQEPG-----CTWGLSSCNFAFCTLCLTYHGVSFCKVTAEKMLDRLNEYLQADEA 383  
 QY 392 -----RVDERAAEQARWEASKEITIKTTKPCPRCHVPVEKNGGCMHKCPQPCRLE 444  
 DB 384 NKRLLDQRYGRVYQKALEMESKEWLEKNSKSCPCGTPLEKLDGCMKMTG--TGCWQY 441  
 QY 445 WCMNCGEMNRVCMG 459  
 DB 442 FCW-----ICWG 448

RESULT 10

US-09-248-796A-15410  
 Sequence 15410, Application US/09248796A  
 Patent No. 6747137  
 GENERAL INFORMATION:  
 APPLICANT: Kelch Weinstein et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
 FILE REFERENCE: 107196.132  
 CURRENT APPLICATION NUMBER: US/09/248,796A  
 CURRENT FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: US 60/074,725  
 PRIOR FILING DATE: 1998-02-13  
 PRIOR APPLICATION NUMBER: US 60/096,409  
 PRIOR FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 28208  
 SEQ ID NO 15410  
 LENGTH: 506  
 TYPE: PRT  
 ORGANISM: Candida albicans  
 US-09-248-796A-15410

Query Match 7.4%; Score 191; DB 4; Length 506;  
 Best Local Similarity 23.6%; Pred. No. 2,7e-10;  
 Matches 79; Conservative 48; Mismatches 118; Indels 90; Gaps 17;

QY 166 CSTCQATVTLTQGSQWMDVLIPLNMSGECSPHCPTSAEFPKCGAHPISDKETPVA 225  
 DB 207 CQIC-QKTL---KGLHC-----TKPDSGCHVFQNNCLAEYFESG----- 241  
 QY 226 LHLIATNSNITCTITDVRSPVLVFOQNSRVITCLDCPHLYCVTLNROQVHPDQOLQY 285  
 DB 242 -----IESGDIDKVNCPD-----FECTKKYVDKKKKF-MELETWVHIDKKVKDMLNQL 288  
 QY 286 SLPCVAGCNSLIKELEHFRILGEEQYNYO--QYGAECVLQMGVLPFRPGCGAGLLP 343  
 DB 289 LIPST--PLAMITKLKSPSLVDRTYVNLFKSQYEWIGNLPLNRLVKCPRGCDDEVIR 345  
 QY 344 EP-DQRKVTCEGNGLGCGPAFCRECKEAYHE-----GEGSAVFEASGTTTQAY--- 391  
 DB 346 EDIDELVVC-----KCKYAFQNDCKSKYHARFKLCLKINENDKYLGIPIEDLEAYPL 400  
 QY 392 -----RVDERAEOARWEAASKETI--KTTKPCPRCHVPEKNGCGM 432  
 DB 401 PPDSDYDKKIMNAKYGRKRID-PAIEEYQMDOLFQMMRERKTVVQCPGCVATEKSEGN 459  
 QY 433 HMKCPQPCRLMECMNGCEMNRVCMG---DHMFD 464  
 DB 460 KMKC-SLCKTDFCFNGSK-----IGNNHDFVD 487

RESULT 11  
 US-08-398-008A-2  
 Sequence 2, Application US/08398008A  
 Patent No. 5665388  
 GENERAL INFORMATION:  
 APPLICANT: Kornbluth, Jacki  
 TITLE OF INVENTION: DNA Encoding Natural Killer Lytic Associated  
 TITLE OF INVENTION: Protein  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Glibbrech & Adler, P.C.  
 STREET: 8011 Candle Lane  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: USA  
 ZIP: 77071  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
 COMPUTER: MACINTOSH IICI  
 OPERATING SYSTEM: Macintosh  
 SOFTWARE: Microsoft Word 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/398,008A  
 FILING DATE: March 2, 1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/126,501  
 FILING DATE: 24-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Adler, Dr. Benjamin Aaron  
 REGISTRATION NUMBER: 35,423  
 REFERENCE/DOCKET NUMBER: D5705CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (713) 777-2321  
 TELEFAX: (713) 777-6908  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 587  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULAR TYPE: Protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-398-008A-2

Query Match 6.4%; Score 167; DB 1; Length 587;  
 Best Local Similarity 24.2%; Pred. No. 8.4e-08;  
 Matches 80; Conservative 24; Mismatches 114; Indels 112; Gaps 20;

QY 178 QPSCWMDVLIPLNMSGECSPHCPTSAEFPKCGAHPISDKETPVALHLIATNSRNI 236  
 DB 73 QGPP-----PEALPAE-----PAAEAEAAAAAGAFPGDDE--AAEGGPGGEV 117  
 QY 237 TCITC-----TDVRSPLVFOQNSRVITCLDCPHLY-----CVTRLND-- 274  
 DB 118 ECPCLIVRLPERRARLL---SCPHRSRCDLRHYLRLEISSRVPISCPESERLNPMD 174  
 QY 275 -RQFVHPDQOLGYSLPVAGCNSLIKELEHFRILGEEQYNYOQYGAECVLQMGVLCF 333  
 DB 175 IRLILADPPL-----MKY-----EEFIVRRYILASDPDC-----RWCP 207  
 QY 334 RPGCGAGLL-----PEPDQRKVTCEGNGLGCGPAFCRECKEAYHEGEGSAVFEASGT 386  
 DB 208 APDCGAVLAIVGACSCP-----KLTCERE--GCOTEFYHKQIWHNPQTCM--ARQ 257  
 QY 387 TQAVRVDERAEOARWEAASKETIKTTKPCPRCHVPEK--NGCMMKCPQPCRL 444  
 DB 258 RAQTLARVTKHTSGLSY---GEGSGDDIKPCPRCSAYIILKNDSCHNMTG--AVCGE 312  
 QY 445 WCMNC-----GC-----EMNR 455  
 DB 313 FCLMKKEISDLHYLSPSGCTFWGKKPMWR 342

RESULT 12  
 US-08-893-333-2  
 Sequence 2, Application US/08893333A  
 Patent No. 5981705  
 GENERAL INFORMATION:  
 APPLICANT: Kornbluth, Jacki  
 TITLE OF INVENTION: DNA Encoding Natural Killer Lytic Associated Protein  
 FILE REFERENCE: D5705CIP/D  
 CURRENT APPLICATION NUMBER: US/08/893,333A  
 CURRENT FILING DATE: 1997-07-16  
 NUMBER OF SEQ ID NOS: 17  
 SEQ ID NO 2  
 LENGTH: 587  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: Amino acid sequence of Natural Killer Lytic  
 OTHER INFORMATION: Associated Protein encoded by nucleotides  
 OTHER INFORMATION: 190 to 1953 of Sequence ID. No. 5981705  
 Patent No. 5981705  
 US-08-893-333-2

Query Match 6.4%; Score 167; DB 2; Length 587;

Best Local Similarity 24.2%; Pred. No. 8.4e-08;  
 Matches 80; Conservative 24; Mismatches 114; Indels 112; Gaps 20;

QY 178 QPSCWMDVLIPLNMSGECSPHCPTSAEFPKCGAHPISDKETPVALHLIATNSRNI 236  
 DB 73 QGPP-----PEALPAE-----PAAEAEAAAAAGAFPGDDE--AAEGGPGGEV 117  
 QY 237 TCITC-----TDVRSPLVFOQNSRVITCLDCPHLY-----CVTRLND-- 274  
 DB 118 ECPCLIVRLPERRARLL---SCPHRSRCDLRHYLRLEISSRVPISCPESERLNPMD 174

QY 275 -RQFVHDPOLGSLFPCVAGCENSLIKELHFRILGEEQYNRQOQYGAECVLQMGVLCG 333  
 Db 175 IRLIADPPL-----MHXY-----EEFIRRIYASPPD-----RWCG 207  
 QY 334 RFGCAGL-----PEPDQRKTYGCGNGLGGFAPFCECKEYAHGECNAVFASGT 386  
 Db 208 APDGGYAVIANGCASC-----KLTICRE-----GGTFEFCYCHQIWHNPQTCM--ARQ 257  
 QY 387 TTAQVYDERAAEQARWEAASKETIKTKTPRCPCYHVEK--NGCGMKCPQPCRL 444  
 Db 258 RAQTLARVTKHTSGLSY---GQSGSPDDIPCRCSAYIIKMDGSCNMHTC--AVCGCE 312  
 QY 445 WCNVC-----GC-----EMNR 455  
 Db 313 FCHLCMKRISDHIYLSFGCTFWGKKRWSR 342

RESULT 13  
 US-09-538-092-598  
 ; Sequence 598, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glot, Loic  
 ; APPLICANT: Mansfield, Traci A.  
 ; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 ; FILE REFERENCE: 15966-542  
 ; CURRENT APPLICATION NUMBER: US/09/538,092  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: 60/127,352  
 ; PRIOR FILING DATE: 1999-04-01  
 ; PRIOR APPLICATION NUMBER: 60/178,965  
 ; NUMBER OF SEQ ID NOS: 2000-02-01  
 ; SOFTWARE: CuraPatSeqFormatter Version 0.9  
 ; SEQ ID NO 598  
 ; LENGTH: 464  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (0)...(0)  
 ; OTHER INFORMATION: Polypeptide Accession Number YNL068W  
 US-09-538-092-598

Query Match 5.7%; Score 149; DB 4; Length 464;  
 Best Local Similarity 21.5%; Pred. No. 3.8e-06;  
 Matches 59; Conservative 33; Mismatches 105; Indels 78; Gaps 13;  
 QY 235 NITCITCTDVRSPVLPVQ-----CNSRHVCLDCCFHLVCVTRINDROF--VHDPOLGYS- 286  
 Db 177 NYHCICMEKMEKGYRMKLPCEANAVEHYLCRCGCAKSYFTAMIQENRISSVACPCCEYKE 236  
 QY 287 -----LPCVACGPNLSIKELHFRILGEEQYNRQOQYGAECVLQMGVLCG 329  
 Db 237 LKLEDFKSYKMKLKAFTPLI---PVSPLKE-----VIDTELCEYEKKFFYNQATRIISK 288  
 QY 330 -----VLCPRGCGAGLLEP-DQRKVTCEGNGLGGFAPFCECKEYAH-EGECNAV 380  
 Db 289 YCPYACVTCR--CDSWCTKEDLDLDMIQQ-----KGFVFCGFDCLAHMGNYNKCKGK 341  
 QY 381 FEASGTTTQAY-----RVDERAAEQARWEAASKETIKTKT 415  
 Db 342 VSLSTDIIEYLDVTVTSYERKRLKAKYGRVLELVNDYLAEMKDLAIKE--GSNL 399  
 QY 416 KPCPRCHVPERKNGGCMHKKCPQPCRLIEWCWNCGCEMNRVCMGDH 461  
 Db 400 QRCPCKCVVERSEGCNKKKC--EVCCTLPFCFICG 432

RESULT 14  
 US-09-248-796A-16471  
 ; Sequence 16471, Application US/09248796A  
 ; Patent No. 6747137

GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstock et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248,796A  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074,725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,409  
 ; NUMBER OF SEQ ID NOS: 1998-08-13  
 ; SEQ ID NO 16471  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Candida albicans  
 US-09-248-796A-16471

Query Match 5.4%; Score 141; DB 4; Length 328;  
 Best Local Similarity 30.2%; Pred. No. 1.4e-05;  
 Matches 32; Conservative 15; Mismatches 37; Indels 22; Gaps 5;  
 QY 358 LCGG--FAFCRECKEAYHEGCSAVFEASGTTQAYVDERAAEQARWEAASKETIKTKT 415  
 Db 58 ICGSEHHEFCNCKENHL-FCPC-----WVSRWIKKCNDDSETAHWIDANT 104  
 QY 416 KPCPRCHVPERKNGGCMHKKCPQPCRLIEWCWNCGCEMNRVCMGDH 461  
 Db 105 HSCPKCYSSIERKNGCUNHMC--QKCYEFCWVCLKDMS-----DH 143

RESULT 15  
 US-08-854-764-2  
 ; Sequence 2, Application US/08854764  
 ; Patent No. 6103500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Innes, Michael  
 ; APPLICANT: Creasey, Abba  
 ; TITLE OF INVENTION: Production of Tissue Factor Pathway  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Chiron Corporation  
 ; STREET: 4560 Horton St.  
 ; CITY: Emeryville  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94608  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.308  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/854,764  
 ; FILING DATE: 12-MAY-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/286,530  
 ; FILING DATE: 05-AUG-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Savereide, Paul B.  
 ; REGISTRATION NUMBER: 36,914  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 510-655-3542  
 ; TELEFAX: 510-655-3542  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 352 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein

US-08-854-764-2

Query Match 5.1%; Score 133; DB 3; Length 352;

Best Local Similarity 18.3%; Pred. No. 0.0001;  
Matches 70; Conservative 56; Mismatches 116; Indels 140; Gaps 16;

```
QY      1 MIVFVFNSSHGFPVEVDSDTISI FOLKEVAKQGVADQLVIFAGKELRNDWTVQNC 60
DB      1 MGIPTKTLGKTLTEVESDITDNVSKIQDKEGIPPDQRLIFAGKCLBDGRTLSDYN 60
QY      61 LDQGSIVHIVQRPWKKGEMNATGDDPKNAAAGCCEREPQSILTRVDLSVLPGDSVGLA 120
DB      61 IQKESTLHLVLR-----LRGGDSE-----DEHTIITDTLPLPLK----- 97
QY      121 VILHT--DGRKDSPPAGSPAGRSIYNSF-----VY-----CKGPCQRY 157
DB      98 --MHSFCAFKADDPCKAIMKRFENITRQCEBFIYGCCEGNQNFESLBECKMKCTRD 155
QY      158 QPGKLRVQCTCROATLTLTGSPSCWDVLIPIRMSEGCQSPH-CPGTSAEFFFK----- 211
DB      156 NANRI-----IKTTLQGEKDPFCLE-----EDPGICRGYITRYFYNNQTKQ 197
QY      212 -----CGAHPISDKETPVALLHLIATNSRNITCITDVRS 246
DB      198 CERFYGCGCLGMNNPETLECKNICEDGPNGPVDNYGTOLNAVNN---SLTPQSTRV 253
QY      247 PVLVPQCNRRHVICLDCEH--LYCVT-----RLNDRQFVHDPQIGYSLP----- 288
DB      254 PSL-PE-----FHGPSMCLTPADRGICRANENRFYNSVIGKCRPFYSGCGGN 301
QY      289 -----CVAGCPNSLIKEL 301
DB      302 ENNFTSKQECCLRAKCKGFIQRI 323
```

Search completed: February 23, 2005, 14:01:03  
Job time : 27.8801 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:57:26 / Search time 73.0:13 Seconds  
(without alignments)  
2084.158 Million cell updates/sec

Title: US-10-622-817-8  
Perfect score: 2596  
Sequence: 1 MIVFVRFNSSHGPFVEVDSD.....CWNCGCEMNRVCMGDHFDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing filter 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US08B\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596	100.0	465	US-10-473-226-2	Sequence 2, Appl1
2	2596	100.0	465	US-10-776-688-2	Sequence 2, Appl1
3	2596	100.0	465	US-10-839-688-9	Sequence 9, Appl1
4	2584	99.5	465	US-10-473-226-4	Sequence 4, Appl1
5	2405	92.6	437	US-10-776-604-4	Sequence 4, Appl1
6	870	33.5	156	US-09-785-548-4	Sequence 4, Appl1
7	544.5	21.0	386	US-10-239-249-2	Sequence 2, Appl1
8	373	14.4	64	US-10-313-203-10	Sequence 10, Appl1
9	323	12.4	56	US-10-313-203-17	Sequence 17, Appl1
10	263	10.1	46	US-09-864-761-36750	Sequence 36750, A
11	246	9.5	525	US-10-437-963-199787	Sequence 199787, A
12	242.5	9.3	604	US-10-425-114-55407	Sequence 55407, A
13	240.5	9.3	503	US-10-080-608A-13	Sequence 13, Appl1

14	240.5	9.3	503	US-10-370-665-102	Sequence 102, App
15	233.5	9.0	585	US-10-424-559-147673	Sequence 147673, A
16	224.5	8.6	648	US-10-425-114-63379	Sequence 63379, A
17	222	8.6	474	US-09-808-387-42	Sequence 42, Appl1
18	222	8.6	474	US-10-442-754-2	Sequence 2, Appl1
19	222	8.6	474	US-10-408-765A-572	Sequence 572, App
20	222	8.6	477	US-10-264-049-3019	Sequence 3019, App
21	219.5	8.5	1753	US-10-146-473-44	Sequence 44, Appl1
22	219.5	8.5	1753	US-10-276-774-1824	Sequence 1824, App
23	219.5	8.5	2517	US-10-313-203-2	Sequence 2, Appl1
24	215	8.3	32	US-10-313-203-18	Sequence 18, Appl1
25	207.5	8.0	623	US-10-437-963-148905	Sequence 148905, A
26	206.5	8.0	452	US-10-425-114-38443	Sequence 38443, A
27	205.5	7.9	449	US-10-425-114-39657	Sequence 39657, A
28	204	7.9	624	US-10-374-780A-1727	Sequence 1727, App
29	204	7.9	624	US-10-437-963-140793	Sequence 140793, A
30	202.5	7.8	289	US-10-437-963-133385	Sequence 133385, A
31	201	7.7	333	US-10-424-559-217796	Sequence 217796, A
32	200	7.7	593	US-10-225-066A-854	Sequence 854, App
33	200	7.7	593	US-10-225-067-56	Sequence 56, Appl1
34	200	7.7	593	US-10-374-780A-330	Sequence 330, App
35	195.5	7.5	541	US-10-424-559-265215	Sequence 265215, A
36	193	7.4	606	US-10-437-963-111378	Sequence 111378, A
37	192	7.4	354	US-10-425-114-56248	Sequence 56248, A
38	191.5	7.4	540	US-10-437-963-168788	Sequence 168788, A
39	191	7.4	338	US-10-425-114-56467	Sequence 56467, A
40	184	7.1	583	US-10-437-963-124556	Sequence 124556, A
41	182.5	7.0	522	US-10-437-963-184651	Sequence 184651, A
42	180	6.9	250	US-10-437-963-152161	Sequence 152161, A
43	180	6.9	394	US-10-425-114-54725	Sequence 54725, A
44	180	6.9	557	US-10-424-559-212592	Sequence 212592, A
45	177.5	6.8	396	US-10-425-114-40231	Sequence 40231, A

#### ALIGNMENTS

RESULT 1  
US-10-473-226-2  
Sequence 2, Application US/10473226  
GENERAL INFORMATION:  
APPLICANT: Megene A/S  
TITLE OF INVENTION: Means for inhibiting proteolytical processing of Parkin  
FILE REFERENCE: 506-204-WO  
CURRENT FILING DATE: 2003-09-29  
CURRENT APPLICATION NUMBER: US/10/473,226  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: DK PA 2001 00525  
PRIOR FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: US 60/281,286  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 465  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: (1)...()  
OTHER INFORMATION: Native Parkin  
US-10-473-226-2

Query Match 100.0%; Score 2596; DB 16; Length 465;  
Best local similarity 100.0%; Pred. No. 2.2e-220;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIVFVRFNSSHGPFVEVDSDSI FOLKEVAVRGGVAPDQLRVTFAGKELRMDVQNC 60  
DB 1 MIVFVRFNSSHGPFVEVDSDSI FOLKEVAVRGGVAPDQLRVTFAGKELRMDVQNC 60  
QY 61 LDQDSIVHVRPMKRGEMNATGDDPRNNAAGGERPQSLTRVDLSSSVLPQDSVGIA 120

Db 61 LDQOSIVHIYQRPWRKQCEMNATGDDPRNAAGCEREPQSLTRVDSLSSVLPGPSVGLA 120  
Qy 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Db 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Qy 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Db 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Qy 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Db 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Qy 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Db 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Qy 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Db 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Qy 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465  
Db 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465

## RESULT 2

US-10-776-604-2  
; Sequence 2, Application US/10776604  
; Publication No. US2005003385A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimizu, Nobuyoshi  
; APPLICANT: Mizuno, Yoshikuni  
; TITLE OF INVENTION: DAs or Genes Participating in Parkinson's Disease  
; FILE REFERENCE: 0652.211001  
; CURRENT APPLICATION NUMBER: US/10/776,604  
; PRIOR FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: 09/601,844  
; PRIOR FILING DATE: 2000-08-09  
; PRIOR APPLICATION NUMBER: PCT/JP99/00545  
; PRIOR FILING DATE: 1999-02-09  
; PRIOR APPLICATION NUMBER: JP 10/27531  
; PRIOR FILING DATE: 1998-02-09  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-776-604-2

Query Match 100.0%; Score 2596; DB 16; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2,2e-220;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIVFVRNSSHGPFVEVDSSTSIIFOLKEVYAKQGVADQLRVIFAGKELRNDWTQNC 60  
Db 1 MIVFVRNSSHGPFVEVDSSTSIIFOLKEVYAKQGVADQLRVIFAGKELRNDWTQNC 60  
Qy 61 LDQOSIVHIYQRPWRKQCEMNATGDDPRNAAGCEREPQSLTRVDSLSSVLPGPSVGLA 120  
Db 61 LDQOSIVHIYQRPWRKQCEMNATGDDPRNAAGCEREPQSLTRVDSLSSVLPGPSVGLA 120  
Qy 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Db 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Qy 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Db 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Qy 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Db 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300

Db 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Qy 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Db 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Qy 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Db 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Qy 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465  
Db 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465

## RESULT 3

US-10-839-688-9  
; Sequence 9, Application US/10839688  
; Publication No. US20050014173A1  
; GENERAL INFORMATION:  
; APPLICANT: Farber, Matthew J.  
; TITLE OF INVENTION: PARKINSON'S DISEASE MARKERS  
; FILE REFERENCE: 07039-448001  
; CURRENT APPLICATION NUMBER: US/10/839,688  
; PRIOR FILING DATE: 2004-05-05  
; PRIOR APPLICATION NUMBER: US 60/468,832  
; PRIOR FILING DATE: 2003-05-08  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-839-688-9

Query Match 100.0%; Score 2596; DB 17; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2,2e-220;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIVFVRNSSHGPFVEVDSSTSIIFOLKEVYAKQGVADQLRVIFAGKELRNDWTQNC 60  
Db 1 MIVFVRNSSHGPFVEVDSSTSIIFOLKEVYAKQGVADQLRVIFAGKELRNDWTQNC 60  
Qy 61 LDQOSIVHIYQRPWRKQCEMNATGDDPRNAAGCEREPQSLTRVDSLSSVLPGPSVGLA 120  
Db 61 LDQOSIVHIYQRPWRKQCEMNATGDDPRNAAGCEREPQSLTRVDSLSSVLPGPSVGLA 120  
Qy 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Db 121 VILHTDSRKDSPAGSPAGRSIYNSFYVYCKGPCQVOPGKLRVOCSTROATLTLTQGP 180  
Qy 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Db 181 SCMDVLIIPNRMSGECQSPHCPGTSAPFFKCGAHTSDKETPVALLHLIATNSRNTICIT 240  
Qy 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Db 241 CTDVRSPLVIFQCNRSRNVICLDCFHLIYCVTRLNDQFVHDPOLGYSLPCVACGPNLSIKE 300  
Qy 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Db 301 LHHFRILGEEQYRNYQOYGAEBEVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC 360  
Qy 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Db 361 GFAPFCEKEAYHEGCSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKKTTKPCPR 420  
Qy 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465  
Db 421 CHVPEKNGGCMHMKCPOQCRLEWCNMGCEMNRVCMGDHMFV 465

```
RESULT 4
US-10-473-226-4
; Sequence 4, Application US/10473226
; Publication No. US20040198650A1
; GENERAL INFORMATION:
; APPLICANT: NAGENE A/S
; TITLE OF INVENTION: Means for inhibiting proteolytical processing of Parkin
; FILE REFERENCE: 506-204-WO
; CURRENT APPLICATION NUMBER: US/10/473,226
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: DK PA 2001 00525
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/281,286
; PRIOR FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..(1)
; OTHER INFORMATION: Parkin with a D126E mutation
US-10-473-226-4

Query Match          99.5%; Score 2584; DB 16; Length 465;
Best Local Similarity 99.6%; Pred. No. 2.5e-219;
Matches 463; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIVFRRFNSHGFPEVVDSDTISIFOLKEVAVAKRGVPADQLRVIFAGKELRNDWTQVQNC 60
DB 1 MIVFRRFNSHGFPEVVDSDTISIFOLKEVAVAKRGVPADQLRVIFAGKELRNDWTQVQNC 60
QY 61 LDQOSIVHIVQRPWRKGGEMNATGDDPRNAAAGCCEREPQSLTRVDLSSSVLPGDSVGLA 120
DB 61 LDQOSIVHIVQRPWRKGGEMNATGDDPRNAAAGCCEREPQSLTRVDLSSSVLPGDSVGLA 120
QY 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
DB 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
QY 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
DB 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
QY 181 SCMDVLLIPNRMSGECQSPHCSTSAEFFPKCGAHTSDKETPVALHLIATNSRNITCT 240
DB 181 SCMDVLLIPNRMSGECQSPHCSTSAEFFPKCGAHTSDKETPVALHLIATNSRNITCT 240
QY 241 CTDVASPVLVFQCNSRHVILCLDCFHLVYCTRLNDRQFVHDPOLGYSLPCVAGCPNSLIKE 300
DB 241 CTDVASPVLVFQCNSRHVILCLDCFHLVYCTRLNDRQFVHDPOLGYSLPCVAGCPNSLIKE 300
QY 301 LHHFRILGEBQYNNRYQOYGAEECVLQMGVLCPRPGCGALLPEBPQRYTCGGNGLGC 360
DB 301 LHHFRILGEBQYNNRYQOYGAEECVLQMGVLCPRPGCGALLPEBPQRYTCGGNGLGC 360
QY 361 GFAFCRECKEAYHBECSAVFEASGTTTQAYRVDERAQAQWEAASKETIKKTPCPR 420
DB 361 GFAFCRECKEAYHBECSAVFEASGTTTQAYRVDERAQAQWEAASKETIKKTPCPR 420
QY 421 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEMNRVCMGDHMFV 465
DB 421 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEMNRVCMGDHMFV 465

RESULT 5
US-10-776-604-4
; Sequence 4, Application US/10776604
; Publication No. US2005000385A1
; GENERAL INFORMATION:
; APPLICANT: Shimizu, Nobuyoshi
; APPLICANT: Mizuno, Yoshikuni
; TITLE OF INVENTION: DNAs or Genes Participating in Parkinson's Disease
; FILE REFERENCE: 0652.2110001
; CURRENT APPLICATION NUMBER: US/10/776,604
```

```
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 09/601,844
; PRIOR FILING DATE: 2000-08-09
; PRIOR APPLICATION NUMBER: PCT/JP99/00545
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: JP 10/27531
; PRIOR FILING DATE: 1998-02-09
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-776-604-4

Query Match          92.6%; Score 2405; DB 16; Length 437;
Best Local Similarity 94.0%; Pred. No. 1.5e-203;
Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MIVFRRFNSHGFPEVVDSDTISIFOLKEVAVAKRGVPADQLRVIFAGKELRNDWTQVQNC 60
DB 1 MIVFRRFNSHGFPEVVDSDTISIFOLKEVAVAKRGVPADQLRVIFAGKELRNDWTQVQNC 60
QY 61 LDQOSIVHIVQRPWRKGGEMNATGDDPRNAAAGCCEREPQSLTRVDLSSSVLPGDSVGLA 120
DB 61 LDQOSIVHIVQRPWRKGGEMNATGDDPRNAAAGCCEREPQSLTRVDLSSSVLPGDSVGLA 120
QY 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
DB 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 180
QY 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 178
DB 121 VILHTDSRKDSPASSPAGRSIYNSFYVCKPQCRVQPGKLRVQOSTROATLTLTQGP 178
QY 181 SCMDVLLIPNRMSGECQSPHCSTSAEFFPKCGAHTSDKETPVALHLIATNSRNITCT 240
DB 179 -----EFPFKGAHTSDKETPVALHLIATNSRNITCT 212
QY 241 CTDVASPVLVFQCNSRHVILCLDCFHLVYCTRLNDRQFVHDPOLGYSLPCVAGCPNSLIKE 300
DB 241 CTDVASPVLVFQCNSRHVILCLDCFHLVYCTRLNDRQFVHDPOLGYSLPCVAGCPNSLIKE 272
QY 301 LHHFRILGEBQYNNRYQOYGAEECVLQMGVLCPRPGCGALLPEBPQRYTCGGNGLGC 360
DB 273 LHHFRILGEBQYNNRYQOYGAEECVLQMGVLCPRPGCGALLPEBPQRYTCGGNGLGC 332
QY 361 GFAFCRECKEAYHBECSAVFEASGTTTQAYRVDERAQAQWEAASKETIKKTPCPR 420
DB 333 GFAFCRECKEAYHBECSAVFEASGTTTQAYRVDERAQAQWEAASKETIKKTPCPR 392
QY 421 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEMNRVCMGDHMFV 465
DB 393 CHVPEKNGGCMHMKCPQPCRLIEWCMNGCEMNRVCMGDHMFV 437

RESULT 6
US-09-785-548-4
; Sequence 4, Application US/09785548
; Patent No. US20020155577A1
; GENERAL INFORMATION:
; APPLICANT: AVENTIS PHARMACEUTICALS, INC.
; TITLE OF INVENTION: COMPOSITIONS THAT CAN BE USED FOR REGULATING THE ACTIVITY OF PARK
; FILE REFERENCE: ST00005
; CURRENT APPLICATION NUMBER: US/09/785,548
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-785-548-4

Query Match          33.5%; Score 870; DB 9; Length 156;
Best Local Similarity 99.4%; Pred. No. 1.3e-68;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 135 GSPAGRSIYNSFYVYCKGPCQORVOPGKLRVOCSTCRQATLTLTGOPSCMDVLIIPNMSG 194
| | | | |
Db 1 GSPAGRSIYNSFYVYCKGPCQORVOPGKLRVOCSTCRQATLTLTGOPSCMDVLIIPNMSG 60
QY 195 ECQSPHCHGTSAEPFKCGAHPSTDKETPVALLIATNSRNITCTTVASPVLFQCN 254
| | | | |
Db 61 ECQSPHCHGTSAEPFKCGAHPSTDKETPVALLIATNSRNITCTTVASPVLFQCN 120
QY 255 SRHVICLDCEPHLYCVTRLNDROPVHDPOLGYSLPCV 290
| | | | |
Db 121 SRHVICLDCEPHLYCVTRLNDROPVHDPOLGYSLPCV 156

RESULT 7
US-10-239-249-2
; Sequence 2, Application US/10239249.
; Publication No. US20030177507A1
; GENERAL INFORMATION:
; APPLICANT: HONER, MARIUS
; APPLICANT: LINK, MOLEFANG
; APPLICANT: BAUMEISTER, RALF
; TITLE OF INVENTION: NEMATODES AS MODEL ORGANISMS FOR INVESTIGATING
; TITLE OF INVENTION: NEURODEGENERATIVE DISEASES AND, IN PARTICULAR, PARKINSON'S
; TITLE OF INVENTION: DISEASE, USES AND METHODS FOR FINDING SUBSTANCES AND
; TITLE OF INVENTION: GENES WHICH CAN BE USED IN TREATING SUCH DISEASES, AND
; FILE REFERENCE: 02481.1804-00000
; CURRENT APPLICATION NUMBER: US/10/239,249
; PRIOR APPLICATION NUMBER: PCT/EP01/03214
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 100 14 109.9
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-239-249-2

Query Match 21.0%; Score 544.5; DB 14; Length 386;
Best Local Similarity 28.9%; Pred. No. 2.1e-39;
Matches 129; Conservative 59; Mismatches 166; Indels 93; Gaps 13;

QY 22 SIFOLKEVAVAKRGVADQLRVLEFAGKELRNMTVQNCIDQOSIVHIVQRMKGQEMN 81
| | | | |
Db 29 NIEDLTKDYBKLTETPSDELEVVFQCKLSTIMWDLSTETATQIMLRKFNHENG 88
QY 82 ATCGDDPRNAAGCCEREPOSLTRVDLSSVLPEDSVGLAVIILHTSRKDSPPAGRS 141
| | | | |
Db 89 AT-----TAKITDS-----S 99
QY 142 IYNSFYVYCKGPCQORVOPGKLRVOCSTCRQATLTLTGOPSCMDVLIIPNMSGECQSPHC 201
| | | | |
Db 100 IIGSFYVWCKN-CDVYKRGKLRVYCKCKSTVLVAKSEQNSDVLSKSRIPAVEECCT 158
QY 202 PGTSAEFFFKCGAHPSTDKETPVALLIATNSRNITCTTVASPVLFQCN SRHVICL 261
| | | | |
Db 159 PGLFAEFKFKC---LACNDPAALTHVAGNMQWMECCCDCKEAVITDLGN-HITCQ 212
QY 267 DEFHLYCVTRLNDROPVHDPOLGYSLPC-VAGCPNSLIKEHLHFRILBEOYNRYQYGA 320
| | | | |
Db 213 FCFRDIYLSQERFVFVQPPHGFITFCYPCG-NRVVODVHFMIGQTSYSEYQKAT 271
QY 321 BECV-LQMGVGLCPRPGCAGLLPBP--DQKVTCEGGNGJCGFAFCECKEAVHEGEC 377
| | | | |
Db 272 ERLIANDKGVYCCVAVSCGQSFMEPYDDGSGCP-----DCEFSFKKC---FERNC 322
QY 378 SAVFASGTTTOAYVVDRAAEQARWEAASKETIKTKTPCERCHVPEKNGGCMWCKP 437
| | | | |
Db 323 VQSHEDDLRT-----TIDATRRCPKCHVATERNGGCAHHC- 360
```

```
QY 438 QPOCRLEMCNCGCEMNRYCMGDHWD 464
| | | | |
Db 361 -TSCGMWCFCFKCTEWKCEQMDHWFN 386

RESULT 8
US-10-313-203-10
; Sequence 10, Application US/10313203
; Publication No. US20040029134A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Wei
; APPLICANT: Nikolaev, Anatoly
; TITLE OF INVENTION: P53-ASSOCIATED PARKIN-LIKE CYTOPLASMIC PROTEIN, AND RELATED COMPO
; TITLE OF INVENTION: AND METHODS
; FILE REFERENCE: 68106
; CURRENT APPLICATION NUMBER: US/10/313,203
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-313-203-10

Query Match 14.4%; Score 373; DB 15; Length 64;
Best Local Similarity 100.0%; Pred. No. 3e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 RYQYGAEECVLQMGVGLCPRPGCAGLLPDPDRKVTCEGGANGJCGFAFCECKEAYH 373
| | | | |
Db 1 RYQYGAEECVLQMGVGLCPRPGCAGLLPDPDRKVTCEGGANGJCGFAFCECKEAYH 60
QY 374 EGEC 377
| | | | |
Db 61 EGEC 64

RESULT 9
US-10-313-203-17
; Sequence 17, Application US/10313203
; Publication No. US20040029134A1
; GENERAL INFORMATION:
; APPLICANT: Nikolaev, Anatoly
; APPLICANT: Gu, Wei
; TITLE OF INVENTION: P53-ASSOCIATED PARKIN-LIKE CYTOPLASMIC PROTEIN, AND RELATED COMPO
; TITLE OF INVENTION: AND METHODS
; FILE REFERENCE: 68106
; CURRENT APPLICATION NUMBER: US/10/313,203
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-313-203-17

Query Match 12.4%; Score 323; DB 15; Length 56;
Best Local Similarity 100.0%; Pred. No. 6.7e-21;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 CITCTDVASPVLFQCN SRHVICLDCEPHLYCVTRLNDROPVHDPOLGYSLPCVAGC 293
| | | | |
Db 1 CITCTDVASPVLFQCN SRHVICLDCEPHLYCVTRLNDROPVHDPOLGYSLPCVAGC 56

RESULT 10
US-09-864-761-36750
; Sequence 36750, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
```

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/532,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 36750  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL035697.18  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.68  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.76  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.94  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.63  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.68  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.91  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.77  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.73  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.96  
OTHER INFORMATION: EST\_HUMAN HIT: T78558.1, EVALUATE 2.00e+00  
OTHER INFORMATION: SWISSPROT HIT: Q04833, EVALUATE 9.10e-01  
US-09-864-761-36750

Query Match 10.1%; Score 263; DB 9; Length 46;  
Best Local Similarity 100.0%; Pred. No. 1e-15;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

332 YNRYYQYGAECVLTQMGVLCPRPGGAGILPEPDQRKYTCGGNG 357  
1 YNRYYQYGAECVLTQMGVLCPRPGGAGILPEPDQRKYTCGGNG 46

RESULT 11  
US-10-437-963-199787  
Sequence 199787, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukhacov, Andrey A.  
APPLICANT: Bardazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 199787  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_95320C.1.pep  
US-10-437-963-199787

Query Match 9.5%; Score 246; DB 16; Length 525;  
Best Local Similarity 30.1%; Pred. No. 7.1e-13;  
Matches 71; Conservative 26; Mismatches 93; Indels 46; Gaps 11;

232 NSRNTCTCTCVRBPVLYFGQNSRHVTCIDCFHLYCVRRLNDROPVHPQAGSLPCV- 290  
127 NRRRLTCAICPVDVPTGGRSAGCSHFYVCSWGRGV-----RAVGDARCLSRCPD 180  
291 ACCPNSLITKEHLHFRILGEEQYNYRYYQYGAECVLTQMGV-LCPRPGGAGILPEPDQRK 349  
181 PSCPAVAVRELVDAVADGDR-ERFGWFALRSYVESNAGMRWCPGCGC-----SRA 230  
350 VTCEGNG-----LGGFAFCRBC-KEAYHEGCSAVFEASGTTTQAYRYDERAARD 400  
231 VEFVGGDGESESVFSCGHGLCWRCGEARPVASCRTV-----AKWYKSSSS 281

401 --ARWEASKEITIKTTKCPCHVPEKNGGCMHKCPQPCRLMCMNCGEEN 454  
282 ETATWLLAH-----TRGCPKCRLPTEKNLGCMHNTC-RPCLHFEFCWICLKPMH 329

RESULT 12  
US-10-425-114-55407  
Sequence 55407, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack B  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 55407  
LENGTH: 604  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3051-106-A7\_Flt1.pep

US-10-425-114-55407

Query Match 9.3%; Score 242.5; DB 15; Length 604;  
Best Local Similarity 27.7%; Pred. No. 1.7e-12;  
Matches 70; Conservative 31; Mismatches 97; Indels 55; Gaps 12;

QY 221 ETPALMLIATNSNITCTITDVRSFVLVFOCNSRHVICLDCFHLQVTRLNDQ---- 276  
DB 140 EKVPEON---SNRELTGCIICFEFVPRARVESTACGHPYCSWAGYFSTINDGPECIV 196  
QY 277 -FVADPOLGYSLPCVACGPNLSIKELHFRILGEEQYNRQYGAEECVLQMGV-LCPR 334  
DB 197 LRCDEPS-----CGAAGVQDMIVL-----ASDEDKQKTDYILRSYIEDNKKTKWCPA 245  
QY 335 PGCGAGLPPDPORRYTCGEGNG-----LGGFAFCREC-KEAYHEGECSAVEASGTTT 388  
DB 246 PGC-----EYAVTFDAGSGNYDVSCISYFCMNCIEEAHPVDC-----GTYS 289  
QY 389 QAVRVDERAAEQARWEAASKETIKTKTPCRCHVPEKNGGCMEMKCPQPCRLKWCWN 448  
DB 290 KWLKNSAESENNMW-----ILANSKPCPKCRPIEKNOGCMHMTG-TPCKFEFCWL 341  
QY 449 CGCEMNRCVCMGDH 461  
DB 342 CLGAMS-----DH 349

RESULT 13

US-10-080-608A-13  
; Sequence 13, Application US/10080608A  
; Publication No. US20030198956A1  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/10/080,608A  
; CURRENT FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-080-608A-13

Query Match 9.3%; Score 240.5; DB 14; Length 503;  
Best Local Similarity 27.5%; Pred. No. 2.1e-12;  
Matches 72; Conservative 42; Mismatches 111; Indels 37; Gaps 13;

QY 209 FPKCGAHPISDKETPVALLIATNSRNITCTITDVRSFVLVFOCNSRHVICLDCFHLQY 268  
DB 105 FPKC-AHVINPNATVETAIKQKTSRSGCECEICFSQLEPDSMAGLECGHRFCMPCMHXL 163  
QY 269 VTRINDRQPVNDPOLGYSLPCVA-GCPNLSIKELHFRILGEEQYN-RYQYGAEECVLQ 326  
DB 164 STK-----IVAEGIGQITISCAHGC-DILVDVTVALNLTDAVRVYQOOLITNSFV-E 215  
QY 327 MGVVL-CPRPGCG-AGLPPDPORRYTCGEGNGLGGFAFCRECKEAYHEG-ECSAVEE 382  
DB 216 CNQLRMWCSVDCTAYAVKVPYAEPRRVHCK-----GHVFCFACGEMHNDPVKRWL-- 267  
QY 383 ASGTTQAVRVDERAAEQARWEAASKETIKTKTPCRCHVPEKNGGCMEMKCPQPCRL 442  
DB 268 -----KKWKICDDSETSNWIAAN-----TKCPCRSVTIEKDGCMHNVCKNQCK 315  
QY 443 LEMCMNCGCEMNRCVCMGDH 464  
DB 316 NEFCWCLGSWEP--HGSSWYN 335

RESULT 14

US-10-370-685-102

; Sequence 102, Application US/10370685  
; Publication No. US20030215903A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyman, Paul  
; APPLICANT: Goldberg, Edward  
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements  
; FILE REFERENCE: NANE.P-004  
; CURRENT APPLICATION NUMBER: US/10/370,685  
; CURRENT FILING DATE: 2003-02-21  
; PRIOR APPLICATION NUMBER: 10/080,608  
; NUMBER OF SEQ ID NOS: 159  
; SOFTWARE: Patent In version 3.2  
; SEQ ID NO 102  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-370-685-102

Query Match 9.3%; Score 240.5; DB 15; Length 503;  
Best Local Similarity 27.5%; Pred. No. 2.1e-12;  
Matches 72; Conservative 42; Mismatches 111; Indels 37; Gaps 13;

QY 209 FPKCGAHPISDKETPVALLIATNSRNITCTITDVRSFVLVFOCNSRHVICLDCFHLQY 268  
DB 105 FPKC-AHVINPNATVETAIKQKTSRSGCECEICFSQLEPDSMAGLECGHRFCMPCMHXL 163  
QY 269 VTRINDRQPVNDPOLGYSLPCVA-GCPNLSIKELHFRILGEEQYN-RYQYGAEECVLQ 326  
DB 164 STK-----IVAEGIGQITISCAHGC-DILVDVTVALNLTDAVRVYQOOLITNSFV-E 215  
QY 327 MGVVL-CPRPGCG-AGLPPDPORRYTCGEGNGLGGFAFCRECKEAYHEG-ECSAVEE 382  
DB 216 CNQLRMWCSVDCTAYAVKVPYAEPRRVHCK-----GHVFCFACGEMHNDPVKRWL-- 267  
QY 383 ASGTTQAVRVDERAAEQARWEAASKETIKTKTPCRCHVPEKNGGCMEMKCPQPCRL 442  
DB 268 -----KKWKICDDSETSNWIAAN-----TKCPCRSVTIEKDGCMHNVCKNQCK 315  
QY 443 LEMCMNCGCEMNRCVCMGDH 464  
DB 316 NEFCWCLGSWEP--HGSSWYN 335

RESULT 15

US-10-424-599-147673

; Sequence 147673, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 147673  
; LENGTH: 585  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(585)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10436C.1.pep  
US-10-424-599-147673

Query Match

9.0%; Score 233.5; DB 15; Length 585;

Best Local Similarity 26.5%; Pred. No. 1e-11;  
Matches 67; Conservative 29; Mismatches 102; Indels 55; Gaps 12;

QY	221	ETPVALLH	IATNSR	ITCT	TDVRS	PVLV	FOCNS	RHVI	CLD	CFHL	YCV	TR	LN	D	---	RQ	276
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	120	EXPIVQH	---	PNAREL	TCGIC	CFENT	PRARI	EMAS	CGHP	YRIS	CMWEG	YIST	SIND	GP	CLM		176
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
QY	277	FVHDPOL	GSYL	PCVAG	CPNS	LKEL	HFRI	LGE	BOY	NR	YOY	GA	BE	CVL	---	OMG	335
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	177	LSPDX	---	---	CGA	TGRS	---	---	SRFG	SR	DE	DE	KOK	YAR	YLRL	REY	225
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
QY	336	GGGAGL	PEPD	ORKY	TC	BGGN	---	---	GL	CG	PA	FC	REC	---	KEA	YH	388
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	226	GC	---	---	EYAV	TFDA	GT	GN	YDV	SC	LS	CS	YG	FC	MN	CTE	269
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
QY	389	QAYR	DERA	BA	EQAR	WEA	ASK	ETI	KK	TK	PC	PR	CH	VP	EN	KG	448
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	270	KWIL	KNSA	SE	SE	NN	---	---	ILAN	SK	PC	PK	KR	PI	EN	HG	321
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
QY	449	CGCE	NN	RV	CM	GDH											461
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	322	CVGA	WS	---	---	DH											329
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	

Search completed: February 23, 2005, 14:33:24  
Job time : 75.013 secs

*This Page Blank (uspio)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using bw model

Run on: February 23, 2005, 13:34:50 ; Search time 19.3575 Seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-8

Perfect score: 2596  
Sequence: 1 MIVFVRFNSSHGFPVEVDSD.....CNNCGCEMNRVCMDHMPDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	490	18.9	357	2 T23460	hypothetical prote
2	220	8.5	543	2 H84724	probable ARI-like
3	219.5	8.5	1753	2 T00350	hypothetical prote
4	205	7.9	437	2 T16477	hypothetical prote
5	205	7.9	491	2 H87793	protein C27A12.6 f
6	198.5	7.6	644	2 T02366	hypothetical prote
7	196.5	7.6	514	2 G84724	probable ARI-like
8	195	7.5	320	2 T45909	hypothetical prote
9	193	7.4	542	2 A84725	similar to Atrialne
10	190.5	7.3	436	2 H87793	protein C27A12.8 f
11	188	7.2	1048	2 T31653	hypothetical prote
12	186.5	7.2	324	2 T16983	hypothetical prote
13	182.5	7.0	551	2 S38086	hypothetical prote
14	180.5	6.9	594	2 T04783	hypothetical prote
15	178.5	6.9	497	2 G87793	protein C27A12.7 f
16	178.5	6.9	688	2 B86448	hypothetical prote
17	174.5	6.7	408	2 T47498	hypothetical prote
18	174	6.7	348	2 T47498	hypothetical prote
19	173	6.7	565	2 F84721	probable RING zinc
20	162.5	6.3	498	2 UC5983	protein kinase C-1
21	159	6.1	518	2 E84536	hypothetical prote
22	158.5	6.1	451	2 T16481	hypothetical prote
23	157	6.0	488	2 T29562	hypothetical prote
24	155	6.0	532	2 T04748	hypothetical prote
25	152.5	5.9	816	2 T25555	hypothetical prote
26	149	5.7	464	2 S48329	probable membrane
27	147	5.7	1209	2 T52523	hypothetical prote
28	143.5	5.5	155	2 JH0227	ubiquitin / riboso
29	142.5	5.5	155	2 JH0226	ubiquitin / riboso

30	141.5	5.5	468	2 A84601	Mutator-like trans
31	136	5.2	222	2 T47487	hypothetical prote
32	134.5	5.2	229	1 UQJN1	ubiquitin precursor
33	134	5.2	373	2 E84647	hypothetical prote
34	134	5.2	638	2 T02611	hypothetical prote
35	133	5.1	397	2 H84578	probable RING zinc
36	132.5	5.1	869	1 UC4858	VLDL receptor prec
37	132	5.1	80	2 T28305	ORF MSV144 probabl
38	132	5.1	150	2 T30390	probable ubiquitin
39	132	5.1	154	2 S55242	polyubiquitin 2 -
40	132	5.1	156	2 C86439	protein T19E23.13
41	131	5.0	534	2 S34285	polyubiquitin - to
42	130.5	5.0	305	1 S29853	polyubiquitin 4 -
43	130.5	5.0	356	1 UQJTRC	polyubiquitin / r1
44	130	5.0	155	2 S40240	ubiquitin/ribosoma
45	130	5.0	156	2 T52335	ubiquitin extensio

ALIGNMENTS

RESULT 1

T23460 hypothetical protein K08E3.7 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T23460  
R/MCMurray, A.  
Submitted to the EMBL Data Library, November 1996  
A/Reference number: Z19743  
A/Accession: T23460  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-357 <WTL>  
A/Cross-references: UNIPROT:Q9XUS3; EMBL:Z61568; PDB: CAB04599.1; GSPDB: GN00021; CESP: K08E3  
A/Experimental source: clone K08E3  
C/Genetics:  
A/Gene: CESP:K08E3.7  
A/Map position: 3  
A/Introns: 23/3; 72/2; 218/1; 286/2; 311/2

Query Match	18.9%	Score 490	DB 2	Length 357
Best Local Similarity	26.7%	Pred. No. 1.8e-29		
Matches 119; Conservative 53; Mismatches 154; Indels 120; Gaps 12;				
QY	22	SIFOLKEVYAKKQGVADQLRYTFAGKELRNDMTVONCDLQDSYHIYQRPARKQENW 81		
DB	29	NIEDLTQDVEKLTETPSDELEWFCGKLSKSTIMRDLSTPATQIMLRKFNSHNENG 88		
QY	82	ATCGDDPRNAAGCEREPSLFRVDSVGLAVILHTDSRKDSPPAGS 141		
DB	89	AT-----TAKITTS-----S 99		
QY	142	YNSEFYVYKGPORVQPKLRVQSTCRQATLTLLQGPSQMDVLIPNMSGECQSPHC 201		
DB	100	ILGSFYVWCKN-CDVYKRGKLRVYCKGSGSTSVLVSEFQMSDVLKSKRIPAVCECCT 158		
QY	202	PETSAEFPKCGAHPISDKETPVALLIATNSRNITCTIDYRSVLYLPQCSRIVICL 261		
DB	159	PGIFAEFKFKC-----LACNDPAALTHVRGNQMTECCVCDGKEKVIPLIGCN--HITQ 212		
QY	262	DCFHLYCVTRLNDROPVNDPQGLSLPCVAGCNSLIKELHHPRIAGEOYNNYOOGAE 321		
DB	213	FCFRV-----RVQDVHHPHMGQSYSEYQKATE 243		
QY	322	ECV-LQMGVLCPRPCGAGLLPEP--DQKVTCEGANGIGCFACRECKEAYHEGEC 378		
DB	244	RLIAYDQKGVTCQPNVSCGQSFMEPYDDGSCP-----DCFFSFCRC--FRNCV 294		
QY	379	AVFEASGTTTQAVRVDERAABQARWEASKEITIKTKTKCPGRHVEVKKGCGMNAKCPQ 438		
DB	295	COSEDLTRT-----TIDATTRBCPKCHVATERNGCAHTHC-- 331		

QY 439 POCRLWNCVCGCENNVCMGDMFD 464  
DB 332 TSCGMDCWCFKCKTEKKECCQMDHFN 357

## RESULT 2

H84724

Probable ARI-like RING zinc finger protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: H84724  
R:Lin, X.; Keul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: H84724  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-543 <STO>  
A:Cross-references: UNIPROT:Q9SKC3; GB:AE002093; NID:94887759; PIDN:AAD32295.1; GSPDB:GN  
A:Gene: At2g31770  
A:Map position: 2

Query Match  
Best Local Similarity 8.5%; Score 220; DB 2; Length 543;  
Matches 60; Conservative 33; Mismatches 83; Indels 56; Gaps 11;

QY 235 NITCITCTDVRSFVAVFQCNHRVLCDFHLYCVTRIND-----RQFVNDPQLGYSLPC 289  
DB 125 NIQCGICFESYTRFEEIARVSCGHPYCKTCMAGITTKIEDPGCLRYKCEPPS-----C 178  
QY 290 VAGCPNSLIKEIHLHPILIGEEQYNY--QQGAEBCVLMGVLCPRPAGGAGLL--PE 344  
DB 179 SAAVGMDMIEDYETKV--NEKTSRYILRSYVEGKIKW----CPSPGGVAVVERGGSF 232  
QY 345 PQQRKTCGGNGLCGFAFCRECKEAYHEG--ECSAVFBAISGTTQAYRDERAABQARM 403  
DB 233 SSSYDVSL-----CSYRFCMNCSEDAHSYVDDTV-----SKW 266  
QY 404 -----EASKEIKTKTKRCPCHVAVENKGGCMHMKCPQPCRLEWMCNC 449  
DB 267 IFKQDSEENKMWLANSKPCPECKRPIEKNDGCMHTCSAP--CGHEFCWIC 317

## RESULT 3

T00350

hypothetical protein KIAA0708 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T00350  
R:ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.;  
DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A:Reference number: Z14142; MUID:98403880; PMID:9734811  
A:Accession: T00350  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1753 <ISH>  
A:Cross-references: UNIPROT:O81WT3; EMBL:AB014608; NID:93327229; PIDN:BA031683.1; PID:G3  
A:Experimental source: brain  
A:Genetics:  
A:Note: KIAA0708

Query Match  
Best Local Similarity 8.5%; Score 219.5; DB 2; Length 1753;  
Matches 125; Conservative 57; Mismatches 181; Indels 221; Gaps 32;

QY 1 MIVVRPRSSSGFPEVY--DSDTS-----IFQIK 27  
DB 1012 MWLLIKFNQTEEVSVETLKDSDLSPELLQLALVPLTSGNPLTLHGGDFPHGGVLRILH 1071

QY 28 EVAKRQG-----VPADQLRVIFAGKELRNDTV-----QNCDDLPQSI 66

DB 1072 BPGQRSEALMLIPQAYINVERDEKRTLEQRNLLSLVRLKHAHGEKXHLHDQ--L 1129

QY 67 VHIYQRPWRKQENMATGDDPRNAAGCCEBQSLTRVDSVSVLPGSVGLAVIHLTD 126

DB 1130 VCLVLEMMQKQPNPPTLG--HTVAGV-----ACTSIDVLSCL--HLGGQYV---- 1175

QY 127 SRKDSPP-----AGSPAGSIYNSRYTYCKPCQRPVQPKLRYQCSYCRQATLTITGCP 180

DB 1176 KRDRDRQILMAYAPRPMGRCGQADVPCGSGSETSKRPEAV-----ATLASQLP 1228

QY 181 SCMDVULIPNRMSGECQSPHCPGSAEFPFK--CGAPITDKETPVVLAHLIATNS--- 233

DB 1229 -----AGRTMSPO-----EVEGLMQYTRQVETLNLSPDVAQHLSHWGAE 1272

QY 234 -----NITCITCTDVRSFV-----LVFQCNHRV 259

DB 1273 QLLQSYSEDPBPLLAAGLCVHQAAVPRDPDHPVCV--SPLGCDLPLSLC--CMHYC 1328

QY 260 CLDCHLYCVTRINDRQFVNDPQLGYSLPC--VAGCP--NSLIKEIHLHPILIGEEQYNY 314

DB 1329 CKSCWNEVILTRT-----EQNLVINTCPIADCPAPGATIRAI-----VSSPEVLSK 1377

QY 315 YQO-----YGAEECVLMGVLCPRP--GCGAGLLPBDQRKTYCEGANGLG-----CGFA 363

DB 1378 YEKALLRGY--VESC--SNLTWCTNPQCD-----RLICRQGLGCGITCKCKGMA 1423

QY 364 FCBECK--EAYHEGCGAVFEASGTTQAYRVERAABQARM-----EASKE 409

DB 1424 SCFNCSEPEAHYPASC-----GHNSSVDVGGQYGYDGMSVDAOSKH 1463

QY 410 TIKKTRPCRCVAVPEKNGCMHMKCPQPCRLEWMCNCGEW 453

DB 1464 LAKLISRCBSQAPLEKNEBGLHMTG--AKCNHGFWRCLKSW 1505

## RESULT 4

T16477

hypothetical protein F56D2.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T16477  
R:Du, Z.  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid F56D2.  
A:Reference number: Z18519  
A:Accession: T16477  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-437 <DUZ>  
A:Cross-references: UNIPROT:Q20871; EMBL:U13644; NID:9532100; PID:91945502; PIDN:AAB5268-  
A:Experimental source: strain Bristol N2; clone F56D2  
C:Genetics:  
A:Gene: CESP:F56D2.5  
A:Map position: 3  
A:Intons: 41/3; 134/3; 185/3; 223/3; 254/2; 287/2; 414/3

Query Match  
Best Local Similarity 7.9%; Score 205; DB 2; Length 437;  
Matches 104; Conservative 66; Mismatches 169; Indels 176; Gaps 28;

QY 18 DSDTSIFQLEKVVAKRQGVADQLRVIFAGKEL--RNDWYQNCDDLPQSIHIVQRPWR 75  
DB 3 DRDLQIYEL-----BALESVLRKKLAKSSWSMDKNAEI--QGIIEV----- 42

QY 76 KQGEEMATGDDPRNAAGCCEBQSLTRVDSVSVLPGSVGLAVIHLTDRKDSPPAG 135

DB 43 -----GFDNLVPTVYITIGTSDSGQFHLPLDILP-----PIRLKHLNENDPTYS 88

QY 136 SPAGRSIYNSFYVYCKGFCQVRGKLRVOCSTGRQATLTLTGSPGCMDDVLIPNMSG 195

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:25:39 ; Search time 102.258 Seconds  
(without alignment)  
1758.725 Million cell updates/sec

Title: US-10-622-817-9  
Perfect score: 2596  
Sequence: 1 MIVFRFNSHGFPVEVDS.....CWNCGCEWNRVCMGDHFDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq16Dec04:\*  
1: geneeqp19808:\*  
2: geneeqp19908:\*  
3: geneeqp20008:\*  
4: geneeqp20018:\*  
5: geneeqp20028:\*  
6: geneeqp20038:\*  
7: geneeqp20038:\*  
8: geneeqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2596	100.0	465	2	AAV32501 Human par
2	2596	100.0	465	6	AAE30800 Human par
3	2596	100.0	465	6	ABO07156 Human p53
4	2584	99.5	465	6	AAE30801 Human par
5	2405	92.6	437	2	AAV32502 Human par
6	2405	92.6	437	6	ABO07157 Human p53
7	2170.5	83.6	464	4	AAE67517 Amino aci
8	2165.5	83.4	464	4	AAE67531 Amino aci
9	2165.5	83.4	464	4	AAE67532 Amino aci
10	2087.5	80.4	451	4	AAE67533 Amino aci
11	1696.5	65.4	316	6	ABO07158 Human p53
12	1679.5	64.7	344	4	AAE67521 Amino aci
13	1234.5	47.6	296	4	AAE67526 Amino aci
14	1058.5	40.8	468	4	ABE65114 Drosophila
15	1002	38.6	250	4	AAE67519 Amino aci
16	997	38.4	262	4	AAE67518 Amino aci
17	870	33.5	156	4	AAE67213 Amino aci
18	671.5	25.9	183	4	AAE67525 Amino aci
19	671.5	25.9	194	4	AAE67524 Amino aci
20	544.5	21.0	386	4	AAE6951 C.elegans
21	506	19.5	153	4	AAE67523 Amino aci
22	335	12.9	77	4	AAE67529 Amino aci
23	263	10.1	46	4	AAE17055 Peptide #
24	263	10.1	46	4	ABE36052 Peptide #
25	263	10.1	46	4	AAE29548 Peptide #

26	263	10.1	46	4	ABE30873 Peptide #
27	263	10.1	46	4	ABE21452 Protein #
28	263	10.1	46	4	AAE69222 Human bon
29	263	10.1	46	4	AAE56839 Human bra
30	263	10.1	46	4	ABE50898 Human liv
31	263	10.1	46	5	ABE38830 Human pep
32	252	9.7	106	8	ADN96489 Modified
33	251	9.7	63	4	AAE67522 Amino aci
34	251	9.7	105	4	AAE67520 Amino aci
35	251	9.7	105	8	ADN96487 Modified
36	240.5	9.3	503	4	ABE61708 Drosophila
37	240.5	9.3	503	7	ADN67082 Atrialine-1
38	240	9.2	511	4	ABE63665 Drosophila
39	239	9.2	520	8	ADP22562 Sea-squid
40	236.5	9.1	445	3	ADP21034 Human nuc
41	236.5	9.1	445	8	ADN96154 T cell ac
42	236.5	9.1	445	8	ADN96026 T cell ac
43	236.5	9.1	557	3	AAE98059 Human Rtn
44	236.5	9.1	557	8	ADN96028 T cell ac
45	236	9.1	53	4	ABE67528 Amino aci

## ALIGNMENTS

RESULT 1	AAV32501	standard; protein; 465 AA.
ID	AAV32501	standard; protein; 465 AA.
AC	AAV32501;	
XX		
DT	21-OCT-1999	(first entry)
XX		
DE	Human parkin gene variant protein.	
XX		
KW	Parkin's disease related gene; parkin gene; variant; gene therapy.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9940191-A1.	
XX		
PD	12-AUG-1999.	
XX		
PF	09-FEB-1999;	99WO-JP000545.
XX		
PR	09-FEB-1998;	98JP-00027531.
XX		
PA	(SHIM/) SHIMIZU N.	
XX		
PI	(MIZU/) MIZUNO Y.	
XX		
XX	Shimizu N, Mizuno Y;	
DR	WPI; 1999-494295/41.	
XX		
DR	N-PSDB; AAX99923.	
XX		
PT	Gene implicated in the pathology of Parkinson's disease, used for	
PT	treatment of the disease.	
XX		
PS	Claim 1; Page 83-88; 11pp; English.	
XX		
CC	This sequence is encoded by a gene of the invention, and is implicated in	
CC	the pathology of Parkinson's disease. This sequence is a variant of the	
CC	parkin gene found in parkinson's disease patients. The sequences may be	
CC	used for the diagnosis, treatment (including gene therapy) and	
CC	investigation of Parkinson's disease	
XX		
SQ	Sequence 465 AA;	
Query Match	100.0%;	Score 2596; DB 2; Length 465;
Best Local Similarity	100.0%;	Pred. No. 2.3e-219;
Matches	465;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MIVFRFNSHGFPVEVDSITIFQLKEVVAKRGVAPADQLRVIFAGKELRMDWTQVNC	60

```

Db      1 MIVFVRFNSSHGFPVEVDSDTSTIFQLKEVAKRGQVADQLRVIFAGKELRNDMTVQNC
QY      61 LDQOSIVHIVORPMKRGEMNATGDDPRNAAAGCCEREPOSILTRVLDSSSVLPQDSVGLA
Db      61 LDQOSIVHIVORPMKRGEMNATGDDPRNAAAGCCEREPOSILTRVLDSSSVLPQDSVGLA
QY      121 VILHTRSRKSDPPAGSPAGRSIYNSFYVYCKGPCQVRQPKLAVQSTCKQATLTLTQGP
Db      121 VILHTRSRKSDPPAGSPAGRSIYNSFYVYCKGPCQVRQPKLAVQSTCKQATLTLTQGP
QY      181 SCMDVLIIPNRMSEGCSPHCPGTSAPFFKCAHPTSDKETPVALHLIATNSRNTTCT
Db      181 SCMDVLIIPNRMSEGCSPHCPGTSAPFFKCAHPTSDKETPVALHLIATNSRNTTCT
QY      241 CTDVRSPLVLFQCNSSRHVILCLDFHLYCVTRLNDROPVHDPOLGSLPCVACCPNSLIKE
Db      241 CTDVRSPLVLFQCNSSRHVILCLDFHLYCVTRLNDROPVHDPOLGSLPCVACCPNSLIKE
QY      301 LHHFRILGEBQYNRVQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC
Db      301 LHHFRILGEBQYNRVQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC
QY      361 GFAPFCRECKEAYHEGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCR
Db      361 GFAPFCRECKEAYHEGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCR
QY      421 CHVPEKNGGCMHMKCPQPCRLIEWCNCGEWNRVCMGDHMFV
Db      421 CHVPEKNGGCMHMKCPQPCRLIEWCNCGEWNRVCMGDHMFV

```

RESULT 2  
AAE30800  
ID AAE30800 standard; protein; 465 AA.

```

AC      AAE30800;
XX
DT      24-FEB-2003 (first entry)
XX
DE      Human Parkin protein.
XX
KW      Human; Parkin protein; neurological disorder; apoptosis; gene therapy;
KW      ischaemic stroke; Parkinson's disease; Alzheimer's disease; neurotrophic;
KW      transgenic; cerebrotective; neuroprotective; neurotransplantation.
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      Cleavage-site 126
XX
XX      MO300279459-A2.
XX
PD      10-OCT-2002.
XX
PF      02-APR-2002; 2002MO-DK000221.
XX
PR      29-MAR-2001; 2001DK-00000525.
PR      03-APR-2001; 2001US-0281286P.
XX
PA      (NSG-) NSGENE AS.
XX
XX      Jensen PH;
XX
XX      WPI; 2003-046812/04.
XX      N-PSDB; AAD47679.
XX
XX      New isolated nucleic acid sequence encoding a Parkin polypeptide, useful
XX      for treating, preventing or diagnosing neurological disorders, e.g.
XX      Parkinson's disease, Alzheimer's disease or ischemic stroke, and in
XX      screening assays.
XX
PS      Claim 10; Page 69; 71pp; English.

```

```

XX      The invention relates to Parkin protein and its corresponding nucleic
CC      acid sequence. The nucleic acid sequence is useful for altering the
CC      proteolytic processing of Parkin at its potential cleavage site at Asp
CC      126. The invention is used in manufacturing or testing a pharmaceutical
CC      composition for treating and/or preventing a neurological disorder, e.g.
CC      Alzheimer's disease or ischemic stroke. It also used for detecting the
CC      occurrence of proteolytic processing of Parkin at Asp 126 in a sample, in
CC      monitoring a potential disposition for a neurodegenerative disease, and
CC      for treating, preventing and/or diagnosing Parkinson's disease or other
CC      neurodegenerative disorders. The viral vector is used for transforming
CC      neuronal cells in vivo or ex vivo. The invention is used for transforming
CC      neurotransplantation into the CNS of a mammal. It may be used in
CC      screening assays to identify compounds that increase or decrease
CC      apoptosis. It is also used in gene therapy. The present sequence is human
CC      Parkin protein.
XX
SQ      Sequence 465 AA;

```

```

Query Match      100.0%; Score 2596; DB 6; Length 465;
Best Local Similarity 100.0%; Pred. No. 2.3e-219; Indels 0; Gaps 0;
Matches 465; Conservative 0; Mismatches 0;

```

```

QY      1 MIVFVRFNSSHGFPVEVDSDTSTIFQLKEVAKRGQVADQLRVIFAGKELRNDMTVQNC
Db      1 MIVFVRFNSSHGFPVEVDSDTSTIFQLKEVAKRGQVADQLRVIFAGKELRNDMTVQNC
QY      61 LDQOSIVHIVORPMKRGEMNATGDDPRNAAAGCCEREPOSILTRVLDSSSVLPQDSVGLA
Db      61 LDQOSIVHIVORPMKRGEMNATGDDPRNAAAGCCEREPOSILTRVLDSSSVLPQDSVGLA
QY      121 VILHTRSRKSDPPAGSPAGRSIYNSFYVYCKGPCQVRQPKLAVQSTCKQATLTLTQGP
Db      121 VILHTRSRKSDPPAGSPAGRSIYNSFYVYCKGPCQVRQPKLAVQSTCKQATLTLTQGP
QY      181 SCMDVLIIPNRMSEGCSPHCPGTSAPFFKCAHPTSDKETPVALHLIATNSRNTTCT
Db      181 SCMDVLIIPNRMSEGCSPHCPGTSAPFFKCAHPTSDKETPVALHLIATNSRNTTCT
QY      241 CTDVRSPLVLFQCNSSRHVILCLDFHLYCVTRLNDROPVHDPOLGSLPCVACCPNSLIKE
Db      241 CTDVRSPLVLFQCNSSRHVILCLDFHLYCVTRLNDROPVHDPOLGSLPCVACCPNSLIKE
QY      301 LHHFRILGEBQYNRVQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC
Db      301 LHHFRILGEBQYNRVQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGIGC
QY      361 GFAPFCRECKEAYHEGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCR
Db      361 GFAPFCRECKEAYHEGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCR
QY      421 CHVPEKNGGCMHMKCPQPCRLIEWCNCGEWNRVCMGDHMFV
Db      421 CHVPEKNGGCMHMKCPQPCRLIEWCNCGEWNRVCMGDHMFV

```

RESULT 3  
AB007156  
ID AB007156 standard; protein; 465 AA.

```

AC      AB007156;
XX
DT      13-AUG-2003 (first entry)
XX
DE      Human p53 modifying protein, SEQ ID 116.
XX
KW      Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
KW      antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
KW      lung cancer; ovarian cancer; angiogenesis; cell cycle;
KW      apoptotic disorder; cell proliferation disorder.
XX
OS      Homo sapiens.
XX

```

PN WO200299122-A1.  
XX  
XX 12-DEC-2002.  
XX  
XX 03-JUN-2002; 2002WO-US017382.  
XX  
XX 05-JUN-2001; 2001US-0296076P.  
XX 10-OCT-2001; 2001US-032605P.  
XX 15-FEB-2002; 2002US-0357253P.  
XX  
XX (EXEEL-) EXELIXIS INC.  
XX  
XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP,  
PI  
XX  
XX WPI; 2003-156859/15.  
XX  
XX N-PsDB; ACD13332.  
XX  
XX  
XX Identifying modulators of the p53 pathway for use in treating apoptotic  
XX or cell proliferation disorders, comprises screening for agents that  
XX modulate activity of a human ortholog of genes that modify the p53  
XX pathway in Drosophila.

PS Example 2; Page 399-401; 678bp; English.

The invention relates to identifying (M1) a candidate p53 pathway modulating agent, by contacting an assay system comprising a purified HM polypeptide (human orthologue of genes that modify the p53 pathway in Drosophila) or nucleic acid with a test agent under conditions, where but for the presence of the test agent, the system provides a reference activity, and detecting a test agent-biased activity of the assay system. Also included are modulating (M2) a p53 pathway of a cell (comprising contacting a cell defective in p53 function with a candidate modulator CC that specifically binds to a HM polypeptide comprising an HM amino acid sequence, where p53 function is restored), modulating (M3) a p53 pathway in a mammalian cell (comprising contacting the cell with an agent that specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4) a disease in a patient (comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with a probe for HM expression; (c) comparing the results with a control; and (d) determining whether the comparison indicates a likelihood disease). (M1) is useful for identifying modulators of the p53 pathway. A probe for HM expression is useful for diagnosing breast, colon, kidney, lung and ovarian cancer, in a patient, where the cancer has greater than 25 % expression level. CC Modulators identified by (M1) are useful in a variety of diagnostic and therapeutic applications, where disease or disorder prognosis is related to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell proliferation disorders (e.g. cancer). Another two new methods (M2 and M3) are useful for modulating the p53 pathway of a cell, thus restoring the p53 function of the cell, so that the cell undergoes normal proliferation or progression through the cell cycle. (M2) and (M3) are also useful for treating defects in the p53 pathway such as angiogenic, apoptotic or cell proliferation disorders. The present sequence represents a human p53 pathway modifying protein

Sequence 465 AA;

Query Match 100.0%; Score 2596; DB 6; Length 465;

Local identity: 100.00; Eloc: NO: 2:30 210;  
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIVFVRFNSSHGFPVEVDSDTISIFQLKEVAKRGVPAADQLRVI FACKELRNDWTVQNC D 60

Db 1 MIVFRNSSHGFPVEVSDTSIFQLKEVAKRGVPADQLRVI FAGKELRNDWTVQNC D 60

61 LDQDSIVHIVQRPWRKQEMNATGGDDPRNAGGCEREPOSLTRVDI,SSSVLPGDSVGLA 120

Db 61 LDQDSIVHIVQRPWRKQEMNATGGDDPRNAGGCEREQSLTRVDI.SSSVLPGDSVGLA 1200

QY 121 VILHDSRKDSPAGSPAGRSIYNSFYVYCKGPGQRVQPGKLRVQCSJTCRQATLLTQGP 180

Db 121 VILHDSRKDSPAGSPAGRSIYNSFYVYCKGPCQRVQPKLRVQCSITCRQATLLTQGP 1800

Qy 181 SCWDDVLIPNRMSGECQSPHCPGTSAEFFFKCGAHP TSDKETPVVALLIATNSRNITCIT 240

Db	181	SCDDVLLIENRMSGEQSPHCSTSAEFPFGCAHPIPSDKETPVALLHIANNSNITCIT	240
Qy	241	CTVRSRPLYVFCQNSRHVYCLDCFHLVCTRLANDRFPHDQLGYSLPVAVGENSELKE	300
Db	241	CTVRSRPLYVFCQNSRHVYCLDCFHLVCTRLANDRFPHDQLGYSLPVAVGENSELKE	300
Qy	301	LHHFRILTGEBOYNYRQOYGAECVYLQMGVLCPRGCGAGLLPEBDQRVTCBGGNGLQC	360
Db	301	LHHFRILTGEBOYNYRQOYGAECVYLQMGVLCPRGCGAGLLPEBDQRVTCBGGNGLQC	360
Qy	361	GFAFRCCEKBAVHGECSAVFEASGTTQAVRVDERAAEQARWEASKETIKKTTTKPCR	420
Db	361	GFAFRCCEKBAVHGECSAVFEASGTTQAVRVDERAAEQARWEASKETIKKTTTKPCR	420
Qy	421	CHVPEKNGGCGMHMKPCPOQCRLEMCNNGCGENRVCMGDHAFDV	465
Db	421	CHVPEKNGGCGMHMKPCPOQCRLEMCNNGCGENRVCMGDHAFDV	465

AAE30801

XX 5

XX

XX

[illegible]

KW ischaemic stroke; Parkinson's disease; Alzheimer's disease; nootropic

KW mutant; mutein.

OS Homo sapiens.

XX

FT Misc-difference 126

FT	Cleavage-site	126

/Note= "Encoded by CCA"  
FT

PN W0200279459-A2

PD 10-OCT-2002.

PF 02-APR-2002; 2002WO-DK000221.

PR 29-MAR-2001; 2001DK-00000525.

**XX**

XX

XX

DR N-PSDB; AAD47680.

PT New isolated nucleic acid sequence encoding a Parkin polypeptide, useful

PT Parkinson's disease, Alzheimer's disease or ischemic stroke, and in

XX

XX

CC acid sequence. The nucleic acid sequence is useful for altering the

CC 126. The invention is used in manufacturing or testing a pharmaceutical

CC Alzheimer's disease or ischaemic stroke. It also used for detecting the  
CC occurrence of proteolytic processing of parkin at Asp 126 in a sample, in  
CC monitoring a potential disposition for a neurodegenerative disease, and  
CC for treating, preventing and/or diagnosing Parkinson's disease or other  
CC neurodegenerative disorders. The viral vector is used for transforming  
CC neuronal cells in vivo or ex vivo. The invention is useful for  
CC neurocranioplasty into the CNS of a mammal. It may be used in  
CC screening assays to identify compounds that increase or decrease  
CC apoptosis. It is also used in gene therapy. The present sequence is human  
CC parkin mutant protein  
XX

Seq Sequence 465 AA;

Query Match 99.5%; Score 2584; DB 6; Length 465;  
Best Local Similarity 99.6%; Pred. No. 2.6e-218;

Matches 463; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIVFVFNSSHGFPVEVDSDTISIFOLKEVNAKRGVPADQLRVIFAGKELRNMTWQNC 60  
DB 1 MIVFVFNSSHGFPVEVDSDTISIFOLKEVNAKRGVPADQLRVIFAGKELRNMTWQNC 60  
QY 61 LDQOSIVHIVORPMRKQGEENATGDDPRNNAAGCEREPQSLTRVLDSSVLPQDSVGLA 120  
DB 61 LDQOSIVHIVORPMRKQGEENATGDDPRNNAAGCEREPQSLTRVLDSSVLPQDSVGLA 120  
QY 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
DB 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
QY 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
DB 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
QY 181 SCMDVLLIPNRMSEGCSPHCPGTSAEFFPKCAHPTSDKETPVVAHLIATNSRNTTCT 240  
DB 181 SCMDVLLIPNRMSEGCSPHCPGTSAEFFPKCAHPTSDKETPVVAHLIATNSRNTTCT 240  
QY 241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVCYTRLNDRQFVNDPOLGYSLPVACGPNLSIKE 300  
DB 241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVCYTRLNDRQFVNDPOLGYSLPVACGPNLSIKE 300  
QY 301 LHHFRILGEEQYNNRYOYGABECVLOMGVLCPRPGGAGLLPEPDQRKVTCEGNGLGC 360  
DB 301 LHHFRILGEEQYNNRYOYGABECVLOMGVLCPRPGGAGLLPEPDQRKVTCEGNGLGC 360  
QY 361 GFACRECKEAYHGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCPR 420  
DB 361 GFACRECKEAYHGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCPR 420  
QY 421 CHVPEKNGGCMHMKCPQPCRLIEWCNCCEMNRCVCMGDHMFV 465  
DB 421 CHVPEKNGGCMHMKCPQPCRLIEWCNCCEMNRCVCMGDHMFV 465

RESULT 5  
AAV32502 standard; protein; 437 AA.

XX AAV32502;  
AC AAV32502;  
XX

DT 21-OCT-1999 (first entry)

XX Human parkin gene variant protein.

XX Parkinson's disease related gene; parkin gene; variant; gene therapy.

OS Homo sapiens.

XX WO940191-A1.

XX 12-AUG-1999.

XX 09-FEB-1999; 99WO-JP000545.

XX 09-FEB-1998; 98JP-00027531.

XX (SHIMU) SHIMIZU N.

PA (MIZU/) MIZUNO Y.

XX Shimizu N, Mizuno Y;

XX WPI: 1999-494295/41.

XX N-PSDB; AAX99924.

PT Gene implicated in the pathology of Parkinson's disease, used for  
PT treatment of the disease.

XX Claim 1; Page 89-94; 114pp; English.

CC This sequence is encoded by a gene of the invention, and is implicated in  
CC the pathology of Parkinson's disease. This sequence is a variant of the  
CC parkin gene found in parkinson's disease patients. The sequences may be  
CC used for the diagnosis, treatment (including gene therapy) and  
CC investigation of Parkinson's disease  
XX

Seq Sequence 437 AA;

Query Match 92.6%; Score 2405; DB 2; Length 437;  
Best Local Similarity 94.0%; Pred. No. 1.3e-202;

Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MIVFVFNSSHGFPVEVDSDTISIFOLKEVNAKRGVPADQLRVIFAGKELRNMTWQNC 60  
DB 1 MIVFVFNSSHGFPVEVDSDTISIFOLKEVNAKRGVPADQLRVIFAGKELRNMTWQNC 60  
QY 61 LDQOSIVHIVORPMRKQGEENATGDDPRNNAAGCEREPQSLTRVLDSSVLPQDSVGLA 120  
DB 61 LDQOSIVHIVORPMRKQGEENATGDDPRNNAAGCEREPQSLTRVLDSSVLPQDSVGLA 120  
QY 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
DB 121 VILHTDSRSDSPAPSGPAGRSIYNSFYCYCKGPCQRPQGLRVQGSTCRQATLTLTQGP 180  
QY 181 SCMDVLLIPNRMSEGCSPHCPGTSAEFFPKCAHPTSDKETPVVAHLIATNSRNTTCT 240  
DB 181 SCMDVLLIPNRMSEGCSPHCPGTSAEFFPKCAHPTSDKETPVVAHLIATNSRNTTCT 240  
QY 241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVCYTRLNDRQFVNDPOLGYSLPVACGPNLSIKE 300  
DB 241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVCYTRLNDRQFVNDPOLGYSLPVACGPNLSIKE 300  
QY 301 LHHFRILGEEQYNNRYOYGABECVLOMGVLCPRPGGAGLLPEPDQRKVTCEGNGLGC 360  
DB 301 LHHFRILGEEQYNNRYOYGABECVLOMGVLCPRPGGAGLLPEPDQRKVTCEGNGLGC 360  
QY 361 GFACRECKEAYHGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCPR 420  
DB 361 GFACRECKEAYHGECSAVFEASGTTTQAYRVDERAAEQARWEAASKETIKTTKPCPR 420  
QY 421 CHVPEKNGGCMHMKCPQPCRLIEWCNCCEMNRCVCMGDHMFV 465  
DB 421 CHVPEKNGGCMHMKCPQPCRLIEWCNCCEMNRCVCMGDHMFV 465

RESULT 6  
ABO07157

ID ABO07157 standard; protein; 437 AA.

XX ABO07157;

XX 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 117.

XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;  
XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
XX lung cancer; ovarian cancer; angiogenesis; cell cycle;  
XX apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.

XX WO200239122-A1.  
PN  
XX 12-DEC-2002.  
PD  
XX 03-JUN-2002; 2002WO-US017382.  
PF  
XX 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.  
XX  
XX (EXEL-) EXELIXIS INC.  
PA  
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX WPI: 2003-156859/15.  
DR N-PSDB; ACD13333.  
XX  
XX Identifying modulators of the p53 pathway for use in treating apoptotic  
PT or cell proliferation disorders, comprises screening for agents that  
PT modulate activity of a human ortholog of genes that modify the p53  
PT pathway in Drosophila.  
XX  
XX Example 2; Page 401-402; 678bp; English.  
XX  
XX The invention relates to identifying (M1) a candidate p53 pathway  
CC modulating agent, by contacting an assay system comprising a purified HM  
CC polypeptide (human orthologue of genes that modify the p53 pathway in  
CC Drosophila) or nucleic acid with a test agent under conditions, where but  
CC for the presence of the test agent, the system provides a reference  
CC activity, and detecting a test agent-biased activity of the assay system.  
CC Also included are modulating (M2) a p53 pathway of a cell, (comprising  
CC contacting a cell defective in p53 function with a candidate modulator  
CC that specifically binds to a HM polypeptide comprising an HM amino acid  
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
CC in a mammalian cell (comprising contacting the cell with an agent that  
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
CC a disease in a patient (comprising: (a) obtaining a biological sample  
CC from the patient; (b) contacting the sample with a probe for HM  
CC expression; (c) comparing the results with a control; and (d) determining  
CC whether the comparison indicates a likelihood disease). (M1) is useful  
CC for identifying modulators of the p53 pathway. A probe for HM expression  
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
CC in a patient, where the cancer has greater than 25 % expression level.  
CC Modulators identified by (M1) are useful in a variety of diagnostic and  
CC therapeutic applications, where disease or disorder prognosis is related  
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
CC the p53 function of the cell, so that the cell undergoes normal  
CC proliferation or progression through the cell cycle. (M2) and (M3) are  
CC also useful for treating defects in the p53 pathway such as angiogenic,  
CC apoptotic or cell proliferation disorders. The present sequence  
CC represents a human p53 pathway modifying protein  
CC  
XX  
XX Sequence 437 AA;  
SQ  
Query Match 92.6%; Score 2405; DB 6; Length 437;  
Best Local Similarity 94.0%; Pred. No. 1.3e-202;  
Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 181 SCMDVLIIPNMSGSCQSPHCSTAEFFFKCGAHPSTDKETPVALLIATNSRNTCT 240  
DB 179 -----EFPFKGAPHTSDKETPVALLIATNSRNTCT 212  
QY 241 CTDVASPVTVFQCNRSRHVCLDPCFHLXCTTRLNDQFVNDPQGLYLLPCVAGCPNLIIE 300  
DB 213 CTDVASPVTVFQCNRSRHVCLDPCFHLXCTTRLNDQFVNDPQGLYLLPCVAGCPNLIIE 272  
QY 301 LHHFRILGEOYNNRYQOYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGNGLGC 360  
DB 273 LHHFRILGEOYNNRYQOYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGNGLGC 332  
QY 361 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERRAQAWEAASKETIKKTKPCR 420  
DB 333 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERRAQAWEAASKETIKKTKPCR 392  
QY 421 CHVPEKNGGCMHMKCPQPCRLBNCWNGCGEMNRYCMGDHFDV 465  
DB 393 CHVPEKNGGCMHMKCPQPCRLBNCWNGCGEMNRYCMGDHFDV 437

## RESULT 7

AAB67517  
ID AAB67517 standard; protein; 464 AA.  
XX  
XX AAB67517;  
AC  
XX  
DT 29-MAY-2001 (first entry)  
DE  
XX  
XX Amino acid sequence of a murine parkin2 polypeptide.

XX Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
KW brain tumour; head trauma; stroke; vascular irregularity;  
KW metabolic irregularity.  
XX  
XX Mus sp.  
OS  
XX  
XX EP1081225-A1.  
PN  
XX  
XX 07-MAR-2001.  
PD  
XX  
XX 30-AUG-1999; 99EP-00116766.  
PF  
XX  
XX 30-AUG-1999; 99EP-00116766.  
PR  
XX  
XX (BIOF-) BIOFRONTIERA PHARM GMBH.  
PA  
XX  
XX  
PI Luebbert H;  
XX  
XX WPI; 2001-212797/22.  
DR N-PSDB; AAF55244.  
DR  
XX  
XX

PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
PT a transgenic non-human animal as an animal model for neurodegenerative  
PT diseases.  
XX  
XX  
XX Disclosure; Page 17-19; 62pp; English.  
PS  
XX  
XX The present sequence represents a murine parkin2 polypeptide. Mutations  
CC or deletions in the parkin2 gene cause Parkinson's disease in humans. The  
CC human parkin2 gene is located in gene region 6q25.2-27. Parkin2  
CC polypeptides and polynucleotides are useful for analysing  
CC neurodegenerative diseases. They are also useful for testing the efficacy  
CC of the treatment of a neurodegenerative disease such as Parkinson's  
CC disease, Alzheimer's disease, Huntington's disease, amyotrophic lateral  
CC sclerosis, Multi-system atrophy, Wilson's disease, Pick's disease, Prion  
CC disease, and secondary causes inducing Parkinson's syndromes like toxins,  
CC drugs, brain tumours, head trauma, stroke, vascular irregularities or  
CC metabolic irregularities, associated with a less active or non-active  
CC parkin protein  
XX

SQ Sequence 464 AA;  
 Query Match 83.6%; Score 2170.5; DB 4; Length 464;  
 Best Local Similarity 83.4%; Pred. No. 66-182;  
 Matches 388; Conservative 29; Mismatches 47; Indels 1; Gaps 1;

QY 1 MIVFVRENSHGPPVEVDSDTISIFOLKEVAVAKRGVADQLRVIPAGKELRNDMTVQNC 60  
 DB 1 MIVFVRENSHGPPVEVDSDTISIFOLKEVAVAKRGVADQLRVIPAGKELRNDMTVQNC 60  
 QY 61 LEOQSIIVHIVORPMRKSGEMNATGDDPRNMAAGCEREPOSITRVDSLSSVLPEDSVGLA 120  
 DB 61 LEOQSIIVHIVORPMRKSGEMNATGDDPRNMAAGCEREPOSITRVDSLSSVLPEDSVGLA 120  
 QY 121 VILHTDSRKDPPAGSPAGRSIYNSFYCYKGPQORVOPGKLRVQCSCTCROATLTLLQGP 180  
 DB 121 VILHTDSRKDPPAGSPAGRSIYNSFYCYKGPQORVOPGKLRVQCSCTCROATLTLLQGP 180  
 QY 181 SCMDVVLIPNRMSEGCSPHCPGTSIAEFFPKCGAHPITSDKETPVALLIATNSHNITCIT 240  
 DB 181 SCMDVVLIPNRMSEGCSPHCPGTSIAEFFPKCGAHPITSDKETPVALLIATNSHNITCIT 240  
 QY 241 CTDVRSPLVIFQCNRRHVICLDCEFLYCVTRLRNDROFVHDPOLGYSILPCVAGCPSNLIXE 300  
 DB 241 CTDVRSPLVIFQCNRRHVICLDCEFLYCVTRLRNDROFVHDPOLGYSILPCVAGCPSNLIXE 300  
 QY 301 LHHFRILIGEEQYRNYOQYGAEECVLQMGVLCPRPGGAGILPEPDQRKVTCEGNGLGC 360  
 DB 301 LHHFRILIGEEQYRNYOQYGAEECVLQMGVLCPRPGGAGILPEPDQRKVTCEGNGLGC 360  
 QY 361 GFVFCRCCKEAYHGECSAVFEASGTTTOAYRVDERAAEQARWEASKEITIKKTKPCPR 420  
 DB 361 GFVFCRCCKEAYHGECSAVFEASGTTTOAYRVDERAAEQARWEASKEITIKKTKPCPR 420  
 QY 421 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 DB 421 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 QY 465 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 DB 465 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465

RESULT 8  
 AAB67531 standard; protein; 464 AA.  
 AC AAB67531;  
 DT 29-MAY-2001 (first entry)  
 DE Amino acid sequence of a mutated murine parkin2 polypeptide.  
 KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.  
 OS Mus sp.  
 PN EP1081225-A1.  
 PD 07-MAR-2001.  
 PF 30-AUG-1999; 99EP-00116766.  
 PR 30-AUG-1999; 99EP-00116766.  
 PA (BIOF-) BIOFRONTERA PHARM GMBH.  
 PI Luebbert H;  
 DR WPI: 2001-212797/22.  
 DR N-PsDB; AAF55258.  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing

PF a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PS Claim 7; Page 47-49; 62pp; English.  
 CC The present sequence represents a murine parkin2 polypeptide. The  
 CC sequence contains the mutation Lys161Asn. Mutations or deletions in the  
 CC parkin2 gene cause Parkinson's disease in humans. The human parkin2 gene  
 CC is located in gene region 6q25.2-27. Parkin2 polypeptides and  
 CC polynucleotides are useful for analysing neurodegenerative diseases. They  
 CC are also useful for testing the efficacy of the treatment of a  
 CC neurodegenerative disease such as Parkinson's disease, Alzheimer's  
 CC disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-  
 CC system atrophy, Wilson's disease, Pick's disease, Prion disease, and  
 CC secondary causes inducing Parkinson's syndromes like toxins, drugs, brain  
 CC tumours, head trauma, stroke, vascular irregularities or metabolic  
 CC irregularities, associated with a less active or non-active parkin  
 CC protein

SQ Sequence 464 AA;  
 Query Match 83.4%; Score 2165.5; DB 4; Length 464;  
 Best Local Similarity 83.2%; Pred. No. 1.7e-181;  
 Matches 387; Conservative 29; Mismatches 48; Indels 1; Gaps 1;

QY 1 MIVFVRENSHGPPVEVDSDTISIFOLKEVAVAKRGVADQLRVIPAGKELRNDMTVQNC 60  
 DB 1 MIVFVRENSHGPPVEVDSDTISIFOLKEVAVAKRGVADQLRVIPAGKELRNDMTVQNC 60  
 QY 61 LEOQSIIVHIVORPMRKSGEMNATGDDPRNMAAGCEREPOSITRVDSLSSVLPEDSVGLA 120  
 DB 61 LEOQSIIVHIVORPMRKSGEMNATGDDPRNMAAGCEREPOSITRVDSLSSVLPEDSVGLA 120  
 QY 121 VILHTDSRKDPPAGSPAGRSIYNSFYCYKGPQORVOPGKLRVQCSCTCROATLTLLQGP 180  
 DB 121 VILHTDSRKDPPAGSPAGRSIYNSFYCYKGPQORVOPGKLRVQCSCTCROATLTLLQGP 180  
 QY 181 SCMDVVLIPNRMSEGCSPHCPGTSIAEFFPKCGAHPITSDKETPVALLIATNSHNITCIT 240  
 DB 181 SCMDVVLIPNRMSEGCSPHCPGTSIAEFFPKCGAHPITSDKETPVALLIATNSHNITCIT 240  
 QY 241 CTDVRSPLVIFQCNRRHVICLDCEFLYCVTRLRNDROFVHDPOLGYSILPCVAGCPSNLIXE 300  
 DB 241 CTDVRSPLVIFQCNRRHVICLDCEFLYCVTRLRNDROFVHDPOLGYSILPCVAGCPSNLIXE 300  
 QY 301 LHHFRILIGEEQYRNYOQYGAEECVLQMGVLCPRPGGAGILPEPDQRKVTCEGNGLGC 360  
 DB 301 LHHFRILIGEEQYRNYOQYGAEECVLQMGVLCPRPGGAGILPEPDQRKVTCEGNGLGC 360  
 QY 361 GFVFCRCCKEAYHGECSAVFEASGTTTOAYRVDERAAEQARWEASKEITIKKTKPCPR 420  
 DB 361 GFVFCRCCKEAYHGECSAVFEASGTTTOAYRVDERAAEQARWEASKEITIKKTKPCPR 420  
 QY 421 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 DB 421 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 QY 465 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465  
 DB 465 CHVPEKNGGCMHMKCPQPCRLKEMCNNGCEMNRVCMDHMFV 465

RESULT 9  
 AAB67532 standard; protein; 464 AA.  
 AC AAB67532;  
 DT 29-MAY-2001 (first entry)  
 DE Amino acid sequence of a mutated murine parkin2 polypeptide.  
 KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KW brain tumour; head trauma; stroke; vascular irregularity;  
 KW metabolic irregularity.



Query Match	Best Local Similarity	83.2%	Score 2165.5	DB 4	Length 464
Matches 387	Conservative 29	Mismatches 48	Indels 1	Gaps 1	
XX	Mus sp.				
XX	EP1081225-A1.				
XX	07-MAR-2001.				
XX	30-AUG-1999;	99EP-00116766.			
XX	30-AUG-1999;	99EP-00116766.			
XX	30-AUG-1999;	99EP-00116766.			
XX	(BIOF-) BIOFRONTERA PHARM GMBH.				
XX	Luebbert H;				
XX	WPI, 2001-212797/22.				
XX	N-PSDB; AAF55259.				
XX	New polynucleotides encoding mouse parkin2 protein, useful for producing a transgenic non-human animal as an animal model for neurodegenerative diseases.				
XX	Claim 7; Page 49-51; 62pp; English.				
XX	The present sequence represents a murine parkin2 polypeptide. The sequence contains the mutation Thr415Asn. Mutations or deletions in the parkin2 gene cause Parkinson's disease in humans. The human parkin2 gene is located in gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are useful for analyzing neurodegenerative diseases. They are also useful for testing the efficacy of the treatment of a neurodegenerative disease such as Parkinson's disease, Alzheimer's disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease, Pick's disease, prion disease, and secondary causes inducing Parkinson's syndromes like toxins, drugs, brain tumours, head trauma, stroke, vascular irregularities or metabolic irregularities, associated with a less active or non-active parkin protein				
XX	Sequence 464 AA;				
QY	1 MIVFVRPNSSHGPFPEVVDSDTISIFOLKEVVARQGVPAQLRVIFAGKELRNDWTVQNC				60
DB	1 MIVFVRPNSSHGPFPEVVDSDTISILQKEVVARQGVPAQLRVIFAGKELRNDWTVQNC				60
QY	61 LDQGSIVHIVQRPWRKGDENATGDDPNNAAAGCCEREPQSLTRVDLSSSVLPQDSVGIA				120
DB	61 LEOGSIVHIVQRPWRKGDENATGDDPNNAAAGCCEREPQSLTRVDLSSSVLPQDSVGIA				120
QY	121 VILHDSRKSDSPASSPAGRSITYNSFYVYCKPQCRVQVGKLRVQCSTSRQATLTLLTQPR				180
DB	121 VILHDSRKSDSPASSPAGRSITYNSFYVYCKPQCRVQVGKLRVQCSTSRQATLTLLTQPR				179
QY	181 SCMDVLLIPNRRSGECQSPHCPTSGTSAEPFKCGAHPSTPKETPVALHLATATSRNITCT				240
DB	180 SCMDVLLIPNRRSGECQSPHCPTSGTSAEPFKCGAHPSTPKETPVALHLATATSRNITCT				239
QY	241 CTDVASPVLVFQCSNRHVICLDCFHLVCTYTRINDRQFVHDPOLGYSIPCVACCPMSLIKE				300
DB	240 CTDVASPVLVFQCSNRHVICLDCFHLVCTYTRINDRQFVHDPOLGYSIPCVACCPMSLIKE				299
QY	301 LHHFRLIGBEQYNNRYQYVABECVLIOMGVLCPRPCGAGLLPEPDQRKVTGEGNGLGC				360
DB	300 LHHFRLIGBEQYNNRYQYVABECVLIOMGVLCPRPCGAGLLPEPDQRKVTGEGNGLGC				359
QY	361 GFAPFRECKEAHEGECASVFAFGSTTTQAYVDEDAEAQAEWMAASKETTICKTKPCPR				420
DB	360 GFAPFRECKEAHEGECASVFAFGSTTTQAYVDEDAEAQAEWMAASKETTICKTKPCPR				419
QY	421 CHVPEKNGGCMHMKCPOPQRLCWCWNCGCEMNRVCMGMDHFDV				465

[illegible]

```
Db 121 VILDTDSKRDSEANAGPV-KFTYNSFFIYCKGPKCHKQVPGKLRVQCGTCKQATLTLAAGP 179
Qy 181 SCMDVDLIPNRMSCGSPHCPGTSAPFPKCAHPTSDKETPVYALHLIATNSRNTICIT 240
Db 180 SCMDVDLIPNRMSCGSPHCPGTSAPFPKCAHPTSDKETPVYALHLIATNSRNTICIT 239
Qy 241 CTDVASPVLVFQCSNRHVICTDCEHLVCTRLNDROFVHDPOLGYSLPCVAGCPNSLIKE 300
Db 240 CTDVASPVLVFQCSNRHVICTDCEHLVCTRLNDROFVHDPOLGYSLPCVAGCPNSLIKE 299
Qy 301 LHHFRILIGBEOYNRYQYGAEECVLQMGVLCPRPGCAGLLPPDDQKVTCEGNGLGC 360
Db 300 LHHFRILIGBEOYNRYQYGAEECVLQMGVLCPRPGCAGLLPPDDQKVTCEGNGLGC 359
Qy 361 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERAARAEQARWEASKETIKTTTKPCPR 420
Db 360 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERAARAEQARWEASKETIKTTTKPCPR 419
Qy 421 CHVPEKNGGCMHMKCPQPCRLMCMNCGCE 452
Db 420 CNVPEIKNGGCMHMKCPQPCRLMCMNCGCE 451
```

## RESULT 11

```
AB007158
ID AB007158 standard; protein; 316 AA.
```

```
AC AB007158;
DT 13-AUG-2003 (first entry)
```

```
DE Human p53 modifying protein, SEQ ID 118.
```

```
XX Human; p53 modifier; cytosolic; cancer; cytosolic; antiangiogenic;
XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
XX lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX apoptotic disorder; cell proliferation disorder.
```

```
XX Homo sapiens.
```

```
OS W0200299122-A1.
```

```
XX 12-DEC-2002.
```

```
XX 03-JUN-2002; 2002MO-US017382.
```

```
XX 05-JUN-2001; 2001US-0296076P.
```

```
XX 10-OCT-2001; 2001US-028605P.
```

```
XX 15-FEB-2002; 2002US-0357253P.
```

```
XX (EXEL-) EXELIXIS INC.
```

```
XX Friedman L, Plozman GD, Belvin M, Francis-Liang H, Li D, Funke RP;
```

```
XX WPI; 2003-156859/15.
```

```
XX N-PSDB; ACD13334.
```

```
XX Identifying modulators of the p53 pathway for use in treating apoptotic
XX modulate activity of a human ortholog of genes that modify the p53
XX pathway in Drosophila.
```

```
XX Example 2; Page 402-403; 678bp; English.
```

```
XX The invention relates to identifying (M1) a candidate p53 pathway
XX modulating agent, by contacting an assay system comprising a purified HM
XX polypeptide (human ortholog of genes that modify the p53 pathway in
XX Drosophila) or nucleic acid with a test agent under conditions, where but
XX for the presence of the test agent, the system provides a reference
XX activity, and detecting a test agent-biased activity of the assay system.
XX Also included are modulating (M2) a p53 pathway of a cell (comprising
XX contacting a cell defective in p53 function with a candidate modulator
XX that specifically binds to a HM polypeptide comprising an HM amino acid
```

```
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC in a mammalian cell (comprising contacting the cell with an agent that
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein
```

```
SO Sequence 316 AA;
```

```
Query Match 65.4%; Score 1696.5; DB 6; Length 316;
Best Local Similarity 68.0%; Pred. No. 1.8e-140;
Matches 316; Conservative 0; Mismatches 0; Indels 149; Gaps 1;
```

```
Qy 1 MIVFVRNNSHGFPVNDSDTSTFOLKEVNAKRGVAPADQLRVTFAGKELRNDVTQVNC 60
Db 1 MIVFVRNNSHGFPVNDSDTSTFOLKEVNAKRGVAPADQLRVTFAGKELRNDVTQV--- 57
```

```
Qy 61 LDQGSIVHIVPRMRKGEMNATGDDPRNAAGCEREPSLRTVDSLSSVLPDSDVGLA 120
Db 58 ----- 57
```

```
Qy 121 VILHTRSKDSPAGSPAGRSYNSFYVYCKGQCRVQPKLRVQGSTCRQATLTLTQGP 180
Db 58 ----- 57
```

```
Qy 181 SCMDVDLIPNRMSCGSPHCPGTSAPFPKCAHPTSDKETPVYALHLIATNSRNTICIT 240
Db 58 -----EFPKCGAHPTSDKETPVYALHLIATNSRNTICIT 91
```

```
Qy 241 CTDVASPVLVFQCSNRHVICTDCEHLVCTRLNDROFVHDPOLGYSLPCVAGCPNSLIKE 300
Db 92 CTDVASPVLVFQCSNRHVICTDCEHLVCTRLNDROFVHDPOLGYSLPCVAGCPNSLIKE 151
```

```
Qy 301 LHHFRILIGBEOYNRYQYGAEECVLQMGVLCPRPGCAGLLPPDDQKVTCEGNGLGC 360
Db 152 LHHFRILIGBEOYNRYQYGAEECVLQMGVLCPRPGCAGLLPPDDQKVTCEGNGLGC 211
```

```
Qy 361 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERAARAEQARWEASKETIKTTTKPCPR 420
Db 212 GFAPCRCKEAYHGECSAVFEASGTTTQAYRVDERAARAEQARWEASKETIKTTTKPCPR 271
```

```
Qy 421 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 272 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

```
Qy 465 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 465
Db 316 CHVPEKNGGCMHMKCPQPCRLMCMNCGCEMNRVCMGDHMFV 316
```

KW	brain tumour; head trauma; stroke; vascular irregularity;
KM	metabolic irregularity.
XX	
OS	Mus sp.
PN	EP1081225-A1.
XX	
PD	07-MAR-2001.
XX	
PP	30-AUG-1999; 99EP-00116766.
XX	
PR	30-AUG-1999; 99EP-00116766.
XX	
PA	(BIOF-) BIOFRONTERA PHARM GMBH.
XX	
PI	Luebbert H;
XX	
DR	WPI: 2001-212797/22.
XX	
DR	N-P8DB; AAF55248.
PT	New polynucleotides encoding mouse parkin2 protein, useful for producing
PT	a transgenic non-human animal as an animal model for neurodegenerative
PT	diseases.
XX	
PS	Claim 7; Page 40-41; 62pp; English.
XX	
CC	The present sequence represents a murine parkin2 polypeptide. The
CC	polynucleotide sequence contains a deletion, leading to a truncated
CC	protein. Mutations or deletions in the parkin2 gene cause Parkinson's
CC	disease in humans. The human parkin2 gene is located in gene region
CC	6q25.2-27. Parkin2 polypeptides and polynucleotides are useful for
CC	analysing neurodegenerative diseases. They are also useful for testing
CC	the efficacy of the treatment of a neurodegenerative disease such as
CC	Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC	amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC	Pick's disease, Prion disease, and secondary causes inducing Parkinson's
CC	syndromes like toxins, drugs, brain tumours, head trauma, stroke,
CC	vascular irregularities or metabolic irregularities, associated with a
CC	less active or non-active parkin protein
XX	
SQ	Sequence 344 AA;
Query Match	64.7%; Score 1679.5; DB 4; Length 344;
Best Local Similarity	66.0%; Pred. No. 6.3e-139;
Matches 307; Conservative	14; Mismatches 23; Indels 121; Gaps 1.
QY	1 MIVPFRFSSHGCFPEVVDSDTSPQLKEVVARQGVPPADQLVIFAGKELRDMDVQNCDD 60
DB	1 MIVPFRFSSSGFPEVVDSDTSLDLKEVVARQGVPPADQLVIFAGKELPHMLTV--- 56
QY	61 LDQGSIVHIVGQPRKQGMENATGDDPRNAAGCGERPQSLTRVDLSGSVLPDGSVGLA 120
DB	57 ----- 56
QY	121 VILHTDSRKSDPPAGSPAGRSITVNSFYVYCKGPCQRPVOPGKLRYVGCSTCRQATLLTGGP 180
DB	57 -----QGP 59
QY	181 SCMDVVLIPNRMSGSCQSPHCCTSAEPFKCGAHPDSDKETPVVLAHLIATSRNITGIT 240
DB	60 SCMDVVLIPNRMSGSCQSPDCCGTAEAPFKCGAHPDSDKDTSLNLTLSRRRSIPCLIA 119
QY	241 CTDVASPVLVFOGCSNRHYVCLDCCFHLVCTYTRANDQGFVNDPOLGYSLPCVACGPNLSLKE 300
DB	120 CTDVASPVLVFOGCSNRHYVCLDCCFHLVCTYTRANDQGFVNDPOLGYSLPCVACGPNLSLKE 179
QY	301 LHHFRILGEBQYNNRYQQYGAEECVLQMGSVLCPRGCGAGLLPEPDQRKVTCEGNGLGC 360
DB	180 LHHFRILGEBQYNNRYQQYGAEECVLQMGSVLCPRGCGAGLLPEPDQRKVTCEGNGLGC 239
QY	361 GFAFRCEKCAVHBEBCSAVFEASGTTTQAVYVNDERAAEQAAWEAASKETTKTKPCPR 420
DB	240 GVFPRDCDEAHHEDDCSLLEPSSGATSGAVYVNDERAAEQAAWEAASKETTKTKTKPCPR 299

Qy	421	CHVPEKNGSGMHKCPQPCQRLTMCNNGGEMRVCMGDHPV	465
Db	300	CNVPEKNGSGMHKCPQPCQCLEKCMNNGGEMNPMACMDHPV	344
RESULT 13			
ID	AAB67526		
AC	AAB67526	standard; protein; 296 AA.	
XX			
XX			
DT	29-MAY-2001	(first entry)	
XX			
DE		Amino acid sequence of a murine truncated parkin2 polypeptide.	
XX			
XX			
KW	Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;		
KW	Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;		
KW	Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;		
KW	brain tumour; head trauma; stroke; vascular irregularity;		
KW	metabolic irregularity.		
XX			
OS	Mus sp.		
XX			
PN	EPI081225-A1.		
XX			
PD	07-MAR-2001.		
XX			
PF	30-AUG-1999; 99EP-00116766.		
XX			
PR	30-AUG-1999; 99EP-00116766.		
XX			
PA	(BIOF-) BIOFRONTIERA PHARM GMBH.		
XX			
PI	Luebbert H;		
XX			
DR	WPI; 2001-212797/22.		
XX			
DR	N-PEDB; AAF55253.		
PT			
PT	New polynucleotides encoding mouse parkin2 protein, useful for producing		
XX	a transgenic non-human animal as an animal model for neurodegenerative		
XX	diseases.		
XX			
PS	Claim 7; Page 44-45; 62pp; English.		
XX			
CC	The present sequence represents a murine parkin2 polypeptide. The		
CC	polynucleotide sequence contains a deletion, leading to a truncated		
CC	protein. Mutations or deletions in the parkin2 gene cause Parkinson's		
CC	disease in humans. The human parkin2 gene is located in gene region		
CC	6q25.2-27. Parkin2 polypeptides and polynucleotides are useful for		
CC	analysing neurodegenerative diseases. They are also useful for testing		
CC	the efficacy of the treatment of a neurodegenerative disease such as		
CC	Parkinson's disease, Alzheimer's disease, Huntington's disease,		
CC	amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,		
CC	Pick's disease, Prion disease, and secondary causes including Parkinson's		
CC	syndromes like toxins, drugs, brain tumours, head trauma, stroke,		
CC	vascular irregularities or metabolic irregularities, associated with a		
XX	less active or non-active parkin protein		
XX			
SQ	Sequence 296 AA;		
Query Match	47.6%;	Score 1234.5;	DB 4;
Best Local Similarity	80.0%;	Pred. No. 7.3e-100;	
Matches 232;	Conservative 20;	Mismatches 37;	Indels 1;
		Gaps 1;	
Qy	1	MIVFVRENSHGPFVEVDSDTSTIQLKVEVAKRGVGPADQLRVIFAGKELANDTVQNC	60
Db	1	MIVFVRENSHGPFVEVDSDTSTIQLKVEVAKRGVGPADQLRVIFAGKELANDTVQNC	60
Qy	61	LDQGSIVHIVRPMRKQGENMAGTGDPRNNAAGCEREPQSLTRVDDSSVLPDGSVGLA	120
Db	61	LEQGSIVHIVRPMRRSHETNAGSGDEQSTSEGSIMESRSLTRVDDSSVLPDGSVGLA	120



CC sclerostis, Multi-system atrophy, Wilson's disease, Prion  
CC disease, and secondary causes inducing Parkinson's syndromes like toxins,  
CC drugs, brain tumours, head trauma, stroke, vascular irregularities or  
CC metabolic irregularities, associated with a less active or non-active  
CC parkin protein

XX  
SQ Sequence 250 AA;

Query Match 38.6%; Score 1002; DB 4; Length 250;  
Best Local Similarity 77.6%; Pred. No. 1,7e-79;  
Matches 190; Conservative 20; Mismatches 35; Indels 0; Gaps 0;

QY 1 MIVFVRFNSSHGPPVEVDSDTISIPQLKEVVAKRGVPPADQLRVIFAGKELRNDWTVQNC 60  
DB 1 MIVFVRFNSSYGPVEVDSDTISIPQLKEVVAKRGVPPADQLRVIFAGKELPNHLTVQNC 60  
QY 61 LDQOSIVHIVQRPWRKGQEMNATGDDPRNAGGCEEPQSILTRVDLSSVLPQDSVGIA 120  
DB 61 LEQOSIVHIVQRPWRSHETNAGDEPQSTSEGSIMESRSLTRVDLSHITLPVDSVGIA 120  
QY 121 VILHTDSRKDSPAGSPAGSIYNSFYVCKGPCORVOPGKLRVQCGTCROATLTLTQGP 180  
DB 121 VILDTDSRKDSZAAKGPAYKPTYNSEFFYCKGPCMKVQPGKLRVQCGTCKQATLTLAQP 180  
QY 181 SCMDVLIIPNRMSGECQSPHCPGTAEPFCKGCAHPTSDKPTPVALLIATNSRNITCIT 240  
DB 181 SCMDVLIIPNRMSGECQSPDCPTRAEPFCKGCAHPTSDKDTSVALLIATNSRISIPCTA 240  
QY 241 CTDVR 245  
DB 241 CTDVR 245

Search completed: February 23, 2005, 13:50:13  
Job time : 103.258 secs

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:36:20 ; Search time 25.8801 Seconds  
(without alignments)  
1341.256 Million cell updates/sec

Title: US-10-622-817-9

Perfect score: 2596  
Sequence: 1 MIVFVRFNSSHGFPEVDS...CNCGCEMNRVCMDHPDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2596	100.0	465	4 US-09-601-844B-2	Sequence 2, Appl 1
2	2596	100.0	465	4 US-09-949-016-6516	Sequence 6516, Ap
3	2405	92.6	437	4 US-09-601-844B-4	Sequence 4, Appl 1
4	2405	92.6	437	4 US-09-949-016-6853	Sequence 6853, Ap
5	1696.5	65.4	316	4 US-09-949-016-6854	Sequence 6854, Ap
6	370.5	14.3	117	4 US-09-270-767-32183	Sequence 32183, A
7	240.5	9.3	403	4 US-09-914-259-13	Sequence 13, Appl 1
8	222	8.6	474	4 US-09-354-221-2	Sequence 2, Appl 1
9	222	8.6	487	4 US-09-949-016-10111	Sequence 10111, A
10	191	7.4	506	4 US-09-248-796A-15410	Sequence 15410, A
11	167	6.4	587	1 US-08-398-008A-2	Sequence 2, Appl 1
12	167	6.4	587	2 US-08-893-333-2	Sequence 2, Appl 1
13	149	5.7	464	4 US-09-538-092-598	Sequence 598, Appl 1
14	141	5.4	328	4 US-09-248-796A-16471	Sequence 16471, A
15	133	5.1	352	4 US-08-854-764-2	Sequence 2, Appl 1
16	133	5.1	352	5 PCT-US95-09377-2	Sequence 2, Appl 1
17	131.5	5.1	1106	4 US-09-949-016-9626	Sequence 9626, Ap
18	130.5	5.0	1242	4 US-09-488-270A-2	Sequence 2, Appl 1
19	128.5	4.9	229	2 US-08-726-306A-23	Sequence 23, Appl 1
20	128.5	4.9	229	3 US-08-840-146-20	Sequence 20, Appl 1
21	128.5	4.9	229	3 US-09-360-220-20	Sequence 20, Appl 1
22	128	4.9	160	4 US-09-370-838-205	Sequence 205, App
23	128	4.9	160	4 US-09-854-133-205	Sequence 205, App
24	127	4.9	2321	4 US-09-230-652-2	Sequence 2, Appl 1
25	126	4.9	237	4 US-09-248-796A-19062	Sequence 19062, A
26	126	4.9	846	1 US-08-149-103-3	Sequence 3, Appl 1
27	126	4.9	846	1 US-08-451-883-3	Sequence 3, Appl 1

28	126	4.9	873	1 US-08-393-734-2	Sequence 2, Appl 1
29	126	4.9	873	3 US-08-894-489-2	Sequence 2, Appl 1
30	126	4.9	904	4 US-09-949-016-9528	Sequence 9528, Ap
31	125.5	4.8	103	2 US-08-771-201-9	Sequence 9, Appl 1
32	125	4.8	76	1 US-08-232-815-2	Sequence 2, Appl 1
33	125	4.8	76	1 US-08-350-906-2	Sequence 2, Appl 1
34	125	4.8	76	1 US-09-601-844B-61	Sequence 61, Appl 1
35	125	4.8	76	5 PCT-US95-04536-2	Sequence 2, Appl 1
36	125	4.8	206	4 US-09-248-796A-19430	Sequence 19430, A
37	125	4.8	1121	1 US-07-789-915A-2	Sequence 2, Appl 1
38	125	4.8	1121	1 US-08-005-005C-2	Sequence 2, Appl 1
39	125	4.8	1121	1 US-08-487-203A-2	Sequence 2, Appl 1
40	124.5	4.8	77	6 5510474-3	Patent No. 5510474
41	124.5	4.8	77	6 5510474-3	Patent No. 5510474
42	124.5	4.8	305	4 US-09-866-153-6	Sequence 6, Appl 1
43	124.5	4.8	305	4 US-09-693-467A-6	Sequence 6, Appl 1
44	124.5	4.8	305	4 US-09-270-976-6	Sequence 6, Appl 1
45	124.5	4.8	381	4 US-09-866-153-9	Sequence 9, Appl 1

## ALIGNMENTS

RESULT 1																																																			
US-09-601-844B-2																																																			
Sequence 2, Application US/09601844B																																																			
Patent No. 6716621																																																			
GENERAL INFORMATION:																																																			
APPLICANT: Shimizu, No. 6716621yoshi																																																			
APPLICANT: Mizuno, Yoshikuni																																																			
TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease																																																			
FILE REFERENCE: 0652.2110000																																																			
CURRENT APPLICATION NUMBER: US/09/601,844B																																																			
CURRENT FILING DATE: 2000-08-09																																																			
PRIOR APPLICATION NUMBER: PCT/J99/00545																																																			
PRIOR FILING DATE: 1999-02-09																																																			
NUMBER OF SEQ ID NOS: 70																																																			
SOFTWARE: PatentIn version 3.1																																																			
SEQ ID NO 2																																																			
LENGTH: 465																																																			
TYPE: PRT																																																			
ORGANISM: Homo sapiens																																																			
US-09-601-844B-2																																																			
Query Match																																																			
Best Local Similarity 100.0%; Score 2596; DB 4; Length 465;																																																			
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;																																																			
QY	1	MIVFVRFNSSHGFPEVDS	DTSI	1	FOLKEVYAKRGVPA	DLRV	IFAGKE	LRND	MTVQ	NC	60																																								
DB	1	MIVFVRFNSSHGFPEVDS	DTSI	1	FOLKEVYAKRGVPA	DLRV	IFAGKE	LRND	MTVQ	NC	60																																								
QY	61	LDQGSIVHIVQRPWRK	QGMN	TGDD	PRNAA	GCERE	POS	LTR	VD	LSS	VL	PGDS	VG	LA	120																																				
DB	61	LDQGSIVHIVQRPWRK	QGMN	TGDD	PRNAA	GCERE	POS	LTR	VD	LSS	VL	PGDS	VG	LA	120																																				
QY	121	VILHDSRSDSP	PASP	PAR	ST	YNS	FY	YV	CK	PC	QV	QV	QV	QV	180																																				
DB	121	VILHDSRSDSP	PASP	PAR	ST	YNS	FY	YV	CK	PC	QV	QV	QV	QV	180																																				
QY	181	SCMDVILIPNR	SGSC	QSP	HC	PG	TS	AE	FP	FK	CA	HT	SK	ET	PV	AL	HL	AT	NS	RN	IT	CT	240																												
DB	181	SCMDVILIPNR	SGSC	QSP	HC	PG	TS	AE	FP	FK	CA	HT	SK	ET	PV	AL	HL	AT	NS	RN	IT	CT	240																												
QY	241	CTDVSPV	LV	FO	CNS	RH	VI	CL	DC	FL	Y	CV	TR	LD	RO	FV	AD	POL	GY	SL	PC	VAC	C	PN	S	L	I	X	300																						
DB	241	CTDVSPV	LV	FO	CNS	RH	VI	CL	DC	FL	Y	CV	TR	LD	RO	FV	AD	POL	GY	SL	PC	VAC	C	PN	S	L	I	X	300																						
QY	301	LHHPF	IL	GE	QV	NR	YQ	QV	GA	E	C	V	OM	G	V	L	C	PR	RC	CG	G	L	L	P	EP	OR	V	T	E	G	G	N	G	L	C	360															
DB	301	LHHPF	IL	GE	QV	NR	YQ	QV	GA	E	C	V	OM	G	V	L	C	PR	RC	CG	G	L	L	P	EP	OR	V	T	E	G	G	N	G	L	C	360															
QY	361	GPA	F	R	E	C	K	E	A	V	H	E	E	C	S	A	V	F	A	S	G	T	T	Q	A	V	H	E	R	A	A	E	Q	A	R	M	E	A	S	K	E	T	I	K	T	T	P	C	B	R	420

```
Db      361  GFACRCEKAYHEGCSAVFEASGTTTQAYRDEAAOAWEAASKETIKTTKPCR 420
Qy      421  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 465
Db      421  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 465
```

## RESULT 2

```
US-09-949-016-6516
; Sequence 6516, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6516
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6516
```

```
Query Match      100.0%; Score 2596; DB 4; Length 465;
Best Local Similarity 100.0%; Pred. No. 8.4e-251;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MIVFVRFNSSHGPFVEVDSDTSIFQLKEVAKRGVPADQLRVIFAGKELRNDWTYQNC 60
Db      1  MIVFVRFNSSHGPFVEVDSDTSIFQLKEVAKRGVPADQLRVIFAGKELRNDWTYQNC 60
Qy      61  LDQOSIVHIVORPMRKQEMNATGDDPRNAAGCGEREPQSLTRVDLSSSVLPDGSVGLA 120
Db      61  LDQOSIVHIVORPMRKQEMNATGDDPRNAAGCGEREPQSLTRVDLSSSVLPDGSVGLA 120
Qy      121  VILHTDSKDSPPAGSPAGRSYNSFYVCKGRCQRPQKRVQCGSTCRQATLTLTGCP 180
Db      121  VILHTDSKDSPPAGSPAGRSYNSFYVCKGRCQRPQKRVQCGSTCRQATLTLTGCP 180
Qy      181  SCWDDVLIIPRMGEGECQSPHCPGTSAEFFPKGAHPTSDKETPVALLIATNSNITCIT 240
Db      181  SCWDDVLIIPRMGEGECQSPHCPGTSAEFFPKGAHPTSDKETPVALLIATNSNITCIT 240
Qy      241  CTDVRSPLVLFQCNRNHYICLDFHLVCTRLNDRQFVHDPOLGYSIPCVAGCPNSLIKE 300
Db      241  CTDVRSPLVLFQCNRNHYICLDFHLVCTRLNDRQFVHDPOLGYSIPCVAGCPNSLIKE 300
Qy      301  LHHFRILGSEQYNNRYQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGLGC 360
Db      301  LHHFRILGSEQYNNRYQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGLGC 360
Qy      361  GFACRCEKAYHEGCSAVFEASGTTTQAYRDEAAOAWEAASKETIKTTKPCR 420
Db      361  GFACRCEKAYHEGCSAVFEASGTTTQAYRDEAAOAWEAASKETIKTTKPCR 420
Qy      421  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 465
Db      421  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 465
```

## RESULT 3

```
US-09-601-844B-4
; Sequence 4, Application US/09601844B
```

```
; Patent No. 6716621
; GENERAL INFORMATION:
; APPLICANT: Shimizu, No. 6716621uyoshi
; APPLICANT: Mizuno, Yoshikuni
; TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease
; FILE REFERENCE: 0652.211000
; CURRENT APPLICATION NUMBER: US/09/601,844B
; PRIOR FILING DATE: 2000-08-09
; PRIOR APPLICATION NUMBER: PCT/JP99/00545
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-601-844B-4
```

```
Query Match      92.6%; Score 2405; DB 4; Length 437;
Best Local Similarity 94.0%; Pred. No. 9.6e-232;
Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

Qy      1  MIVFVRFNSSHGPFVEVDSDTSIFQLKEVAKRGVPADQLRVIFAGKELRNDWTYQNC 60
Db      1  MIVFVRFNSSHGPFVEVDSDTSIFQLKEVAKRGVPADQLRVIFAGKELRNDWTYQNC 60
Qy      61  LDQOSIVHIVORPMRKQEMNATGDDPRNAAGCGEREPQSLTRVDLSSSVLPDGSVGLA 120
Db      61  LDQOSIVHIVORPMRKQEMNATGDDPRNAAGCGEREPQSLTRVDLSSSVLPDGSVGLA 120
Qy      121  VILHTDSKDSPPAGSPAGRSYNSFYVCKGRCQRPQKRVQCGSTCRQATLTLTGCP 180
Db      121  VILHTDSKDSPPAGSPAGRSYNSFYVCKGRCQRPQKRVQCGSTCRQATLTLTGCP 180
Qy      181  SCWDDVLIIPRMGEGECQSPHCPGTSAEFFPKGAHPTSDKETPVALLIATNSNITCIT 240
Db      179  -----BFFPKGAHPTSDKETPVALLIATNSNITCIT 212
Qy      241  CTDVRSPLVLFQCNRNHYICLDFHLVCTRLNDRQFVHDPOLGYSIPCVAGCPNSLIKE 300
Db      213  CTDVRSPLVLFQCNRNHYICLDFHLVCTRLNDRQFVHDPOLGYSIPCVAGCPNSLIKE 272
Qy      301  LHHFRILGSEQYNNRYQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGLGC 360
Db      273  LHHFRILGSEQYNNRYQOYGAEECVLQMGVLCPRPGCAGLLPEPDQRKVTCEGNGLGC 332
Qy      361  GFACRCEKAYHEGCSAVFEASGTTTQAYRDEAAOAWEAASKETIKTTKPCR 420
Db      333  GFACRCEKAYHEGCSAVFEASGTTTQAYRDEAAOAWEAASKETIKTTKPCR 392
Qy      421  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 465
Db      393  CHVEKNKGGMHMKCPQPCRLMEKNCCEMNRVCMGDHMFV 437
```

## RESULT 4

```
US-09-949-016-6853
; Sequence 6853, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
```



SOFTWARE: FaastSeq for windows Version 4.0  
 ; SEQ ID NO 6853  
 ; LENGTH: 437  
 ; TYPE: prt  
 ; ORGANISM: Human  
 US-09-949-016-6853

Query Match 92.6%; Score 2405; DB 4; Length 437;  
 Best Local Similarity 94.0%; Pred. No. 9.6e-232;  
 Matches 437; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MIVFVFNSSHGFPVEVDSITSIFOLKEVAVKQGVADQLRVIFAGKELRNDWTQVQNC 60  
 |||||  
 DB 1 MIVFVFNSSHGFPVEVDSITSIFOLKEVAVKQGVADQLRVIFAGKELRNDWTQVQNC 60  
 QY 61 LDOQSI VHI VQRPWRKGGEMNATGDDPRNNAAGCEREPOSILTRVDLSSSVLPGDSVGLA 120  
 |||||  
 DB 61 LDOQSI VHI VQRPWRKGGEMNATGDDPRNNAAGCEREPOSILTRVDLSSSVLPGDSVGLA 120  
 QY 121 VILHTDSRKDSPAGSPAGRSIYNSFYVCKGPCQRPQKLRVQGSTCQATLTLTQGP 180  
 |||||  
 DB 121 VILHTDSRKDSPAGSPAGRSIYNSFYVCKGPCQRPQKLRVQGSTCQATLTLTQGP 180  
 QY 181 SCMDVLI.PNRMSGCQSPHCPTSAEPFKCGAHPSTDKETPVALHLIATNSRNITCIT 240  
 |||||  
 DB 179 -----BPFKCGAHPSTDKETPVALHLIATNSRNITCIT 212  
 QY 241 CTDVASPVLVFOCNSRHVYCLDCEPHLYCYTRLNDROPVNDPOLGYSLPCVACGPNLSLKE 300  
 |||||  
 DB 241 CTDVASPVLVFOCNSRHVYCLDCEPHLYCYTRLNDROPVNDPOLGYSLPCVACGPNLSLKE 300  
 QY 213 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 465  
 |||||  
 DB 213 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 465  
 QY 301 LHHFRILGEOYNRVQYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGGNGLGC 360  
 |||||  
 DB 273 LHHFRILGEOYNRVQYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGGNGLGC 332  
 QY 361 GFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARWEAASKETIKTKPCPR 420  
 |||||  
 DB 333 GFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARWEAASKETIKTKPCPR 392  
 QY 421 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 465  
 |||||  
 DB 393 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 437

RESULT 5  
 US-09-949-016-6854  
 ; Sequence 6854, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FaastSeq for windows Version 4.0  
 ; SEQ ID NO 6854  
 ; LENGTH: 316  
 ; TYPE: prt  
 ; ORGANISM: Human  
 US-09-949-016-6854

Query Match 65.4%; Score 1696.5; DB 4; Length 316;  
 Best Local Similarity 68.0%; Pred. No. 4.2e-161;  
 Matches 316; Conservative 0; Mismatches 0; Indels 149; Gaps 1;

QY 1 MIVFVFNSSHGFPVEVDSITSIFOLKEVAVKQGVADQLRVIFAGKELRNDWTQVQNC 60  
 |||||  
 DB 1 MIVFVFNSSHGFPVEVDSITSIFOLKEVAVKQGVADQLRVIFAGKELRNDWTQVQNC 57  
 QY 61 LDOQSI VHI VQRPWRKGGEMNATGDDPRNNAAGCEREPOSILTRVDLSSSVLPGDSVGLA 120  
 |||||  
 DB 58 -----  
 QY 181 SCMDVLI.PNRMSGCQSPHCPTSAEPFKCGAHPSTDKETPVALHLIATNSRNITCIT 240  
 |||||  
 DB 58 -----  
 QY 241 CTDVASPVLVFOCNSRHVYCLDCEPHLYCYTRLNDROPVNDPOLGYSLPCVACGPNLSLKE 300  
 |||||  
 DB 92 CTDVASPVLVFOCNSRHVYCLDCEPHLYCYTRLNDROPVNDPOLGYSLPCVACGPNLSLKE 151  
 QY 301 LHHFRILGEOYNRVQYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGGNGLGC 360  
 |||||  
 DB 152 LHHFRILGEOYNRVQYGAEECVLQMGVLCPRPCGAGLLPEPDQKVTCEGGNGLGC 211  
 QY 361 GFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARWEAASKETIKTKPCPR 420  
 |||||  
 DB 212 GFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARWEAASKETIKTKPCPR 271  
 QY 421 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 465  
 |||||  
 DB 272 CHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMFV 316

RESULT 6  
 US-09-270-767-32183  
 ; Sequence 32183, Application US/09270767  
 ; Patent No. 6703491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Homburger et al.  
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 ; FILE REFERENCE: File Reference: 7326-094  
 ; CURRENT APPLICATION NUMBER: US/09/270,767  
 ; CURRENT FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 62517  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 32183  
 ; LENGTH: 117  
 ; TYPE: prt  
 ; ORGANISM: Drosophila melanogaster  
 US-09-270-767-32183

Query Match 14.3%; Score 370.5; DB 4; Length 117;  
 Best Local Similarity 53.3%; Pred. No. 3.8e-29;  
 Matches 64; Conservative 11; Mismatches 40; Indels 5; Gaps 2;

QY 345 PDQRVTECGNGGLCGFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARW 403  
 |||||  
 DB 1 PDQRVTECGNGGLCGFAFCRECKEAYHEGCSAVFEASGTTTQAYRVDERAABQARW 56  
 QY 404 EAASKETIKTKPCPRCHVPEKNGGCMHMKCPQPCRLLEWCNCGCEMNRVCMGDHMF 463  
 |||||  
 DB 57 DEASNVITIKVTPCKRTPTERRDGGCMHMKCTAAGCGFEMCWQCQETMTDCMGAHWF 116

RESULT 7  
 US-09-914-259-13  
 ; Sequence 13, Application US/09914259  
 ; Patent No. 6495336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makowski, Lee  
 ; APPLICANT: Hyman, Paul  
 ; APPLICANT: Williams, Mark  
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

```

; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-914-259-13

```

```

Query Match      9.3%; Score 240.5; DB 4; Length 503;
Best Local Similarity 27.5%; Pred. No. 3e-15;
Matches 72; Conservative 42; Mismatches 111; Indels 37; Gaps 13;

```

```

QY 209 FFKCAHPTSDKETPVALHLIATNSRNITCTCTDVRSPVLVQCNSRHVTCIDCFHLYC 268
DB 105 FFKC-AHVINPFNATEAIKOKTNSQCEBCEI CPSQLPDSMAGLECGHRCFPCMCHEYL 163
QY 269 VTRINDQFVNDPQLGYSLEPCVA-GCPNSLIKEHLHFRILGEOYN-RYQYGAEECVLQ 326
DB 164 STK-----IVASLGQTISCAHGC-DILVDVTVANLVTDARVAVKQULITNSFY-E 215
QY 327 MGVVL--CPRPGCG-AGLPEPDRKVTCEGNGLGGCGAFCECKEAYHEG-EGSAVFE 382
DB 216 CNOILRWCPSVDCTYAVKVPYAEPRVHCK-----CGHVCFACGENMHPVXCRWL-- 267
QY 383 ASGTTQAVYVDBRAEQARWEASKETIKKTKPCPRCHVPEKNKGCMHMKCPOPCRL 442
DB 268 -----KKWIKKCCDDSETSNMTAAN-----TKCPCRCSTVIEIDGCHNMVCKNQNC 315
QY 443 LEKWNCGCEMNRVCGMDHMF 464
DB 316 NEFCWVCLGSEWEP--HGSSWN 335

```

```

RESULT 8
US-09-354-221-2
; Sequence 2, Application US/09354221
; Patent No. 6699714
; GENERAL INFORMATION:
; APPLICANT: Chang, Chawmshang
; TITLE OF INVENTION: Androgen Receptor Coactivators
; FILE REFERENCE: 920920.90011
; CURRENT APPLICATION NUMBER: US/09/354,221
; CURRENT FILING DATE: 1999-07-15
; EARLIER APPLICATION NUMBER: US 60/100,243
; EARLIER FILING DATE: 1998-09-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-354-221-2

```

```

Query Match      8.6%; Score 222; DB 4; Length 474;
Best Local Similarity 27.8%; Pred. No. 2e-13;
Matches 71; Conservative 27; Mismatches 97; Indels 60; Gaps 13;

```

```

QY 232 NSRNITC--ITCTDVRSPVLV-F-QCNSRHVTCIDCFHLYCVTRINDQFVNDPQLGYSLP 288
DB 214 NKRLLDQRYGKRVIAQKALEMESKEWLEKNSKSCPCGTPPIEKLDGCMNMTG--TGCMQY 441
QY 289 CVAGCPN-----SLIKELHFRILGEOYNRYQYGAEECVLQMGVLT-CPRPGCG 338
DB 262 CL-NCEPRKCPSPVATPGQVKEL-----VEALFARYDRLLLOSLLDMADVVCPRPCQ 315
QY 339 AGLPEPDRKVTCEGNGLGGCGAFCECKEAYH-EGESAVFEASGTTQAY----- 391
DB 316 LPVWQEPG-----CTMGICSSCNFAFCTLCRLTYHGVSCKYTAELMDLRNEYILOADEA 370
QY 392 -----RVDERAAEQARWEASKETIKKTKPCPRCHVPEKNKGCMHMKCPOPCRL 444

```

```

DB 371 NKRLLDQRYGKRVIAQKALEMESKEWLEKNSKSCPCGTPPIEKLDGCMNMTG--TGCMQY 428
QY 445 WCMNCGCEMNRVCGM 459
DB 429 FCW-----ICMG 435

```

```

RESULT 9
US-09-949-016-10111
; Sequence 10111, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10111
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10111

```

```

Query Match      8.6%; Score 222; DB 4; Length 487;
Best Local Similarity 27.8%; Pred. No. 2e-13;
Matches 71; Conservative 27; Mismatches 97; Indels 60; Gaps 13;

```

```

QY 232 NSRNITC--ITCTDVRSPVLV-F-QCNSRHVTCIDCFHLYCVTRINDQFVNDPQLGYSLP 288
DB 227 NKRLLDQRYGKRVIAQKALEMESKEWLEKNSKSCPCGTPPIEKLDGCMNMTG--TGCMQY 441
QY 289 CVAGCPN-----SLIKELHFRILGEOYNRYQYGAEECVLQMGVLT-CPRPGCG 338
DB 275 CL-NCEPRKCPSPVATPGQVKEL-----VEALFARYDRLLLOSLLDMADVVCPRPCQ 328
QY 339 AGLPEPDRKVTCEGNGLGGCGAFCECKEAYH-EGESAVFEASGTTQAY----- 391
DB 329 LPVWQEPG-----CTMGICSSCNFAFCTLCRLTYHGVSCKYTAELMDLRNEYILOADEA 383
QY 392 -----RVDERAAEQARWEASKETIKKTKPCPRCHVPEKNKGCMHMKCPOPCRL 444
DB 384 NKRLLDQRYGKRVIAQKALEMESKEWLEKNSKSCPCGTPPIEKLDGCMNMTG--TGCMQY 441
QY 445 WCMNCGCEMNRVCGM 459
DB 442 FCW-----ICMG 448

```

```

RESULT 10
US-09-248-796A-15410
; Sequence 15410, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

```



```
Db      61 LDQOSIVHIVORPMKGGEMNATGDDPRNAAGCEREPOSTLRVDLSSSVLPGBSVGLA 120
Qy      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Db      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Qy      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Db      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Qy      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
Db      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
Qy      301 LHHFRILGEBQYNNRYOQYGAEBECVLOMGVLCPRPGCAGLLPEPDQRKTYCEGANGLCG 360
Db      301 LHHFRILGEBQYNNRYOQYGAEBECVLOMGVLCPRPGCAGLLPEPDQRKTYCEGANGLCG 360
Qy      361 GFAFCRECKEAYHEGSCSAVFEASGTTTOAYRVDERAAEQARWEAASKETIKKTYPCPR 420
Db      361 GFAFCRECKEAYHEGSCSAVFEASGTTTOAYRVDERAAEQARWEAASKETIKKTYPCPR 420
Qy      421 CHVPEKNGGCMHMKCPQPCRLEWCNCGCEMNRYCMGDHMFV 465
Db      421 CHVPEKNGGCMHMKCPQPCRLEWCNCGCEMNRYCMGDHMFV 465
```

```
RESULT 2
US-10-776-604-2
/ Sequence 2, Application US/10776604
/ Publication No. US2005000385A1
/ GENERAL INFORMATION:
/ APPLICANT: Shimizu, Nobuyoshi
/ TITLE OF INVENTION: DMS or Genes Participating in Parkinson's Disease
/ FILE REFERENCE: 0652.2110001
/ CURRENT APPLICATION NUMBER: US/10/776,604
/ PRIOR FILING DATE: 2004-02-12
/ PRIOR APPLICATION NUMBER: 09/601,844
/ PRIOR FILING DATE: 2000-08-09
/ PRIOR APPLICATION NUMBER: PCT/JP99/00545
/ PRIOR FILING DATE: 1999-02-09
/ PRIOR APPLICATION NUMBER: JP 10/27531
/ PRIOR FILING DATE: 1998-02-09
/ NUMBER OF SEQ ID NOS: 70
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 2
/ LENGTH: 465
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-776-604-2
```

```
Query Match      100.0%; Score 2596; DB 16; Length 465;
Best Local Similarity 100.0%; Pred. No. 2.2e-220;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVPADQLRVIFAGKELRNDMTVQNC 60
Db      1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVPADQLRVIFAGKELRNDMTVQNC 60
Qy      61 LDQOSIVHIVORPMKGGEMNATGDDPRNAAGCEREPOSTLRVDLSSSVLPGBSVGLA 120
Db      61 LDQOSIVHIVORPMKGGEMNATGDDPRNAAGCEREPOSTLRVDLSSSVLPGBSVGLA 120
Qy      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Db      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Qy      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Db      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Qy      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
Db      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
```

```
RESULT 3
US-10-839-688-9
/ Sequence 9, Application US/10839688
/ Publication No. US20050014173A1
/ GENERAL INFORMATION:
/ APPLICANT: Farret, Matthew J.
/ TITLE OF INVENTION: PARKINSON'S DISEASE MARKERS
/ FILE REFERENCE: 07039-448001
/ CURRENT APPLICATION NUMBER: US/10/839,688
/ PRIOR FILING DATE: 2004-05-05
/ PRIOR APPLICATION NUMBER: US 60/468,832
/ NUMBER OF SEQ ID NOS: 81
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 9
/ LENGTH: 465
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-839-688-9
```

```
Query Match      100.0%; Score 2596; DB 17; Length 465;
Best Local Similarity 100.0%; Pred. No. 2.2e-220;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVPADQLRVIFAGKELRNDMTVQNC 60
Db      1 MIVFVFNSSHGFPVEVDSTISIFOLKEVAKRGQVPADQLRVIFAGKELRNDMTVQNC 60
Qy      61 LDQOSIVHIVORPMKGGEMNATGDDPRNAAGCEREPOSTLRVDLSSSVLPGBSVGLA 120
Db      61 LDQOSIVHIVORPMKGGEMNATGDDPRNAAGCEREPOSTLRVDLSSSVLPGBSVGLA 120
Qy      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Db      121 VILHTDSRKDSPAPGSPAGRSINYSFYVYCKGPCQORVOPGKLRVQGSTCRQATLTLLTQGP 180
Qy      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Db      181 SCMDVLIIPNMSGECSPHCPGTSAPFEFFKCGAHPSTDKETPVALHLIATNSRNITCIT 240
Qy      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
Db      241 CTDVRSPLVLFQCNRSRHVITCLDCFHLVYCTRLNDROFVHDPOLGSLPCVAGCENSLIKE 300
Qy      301 LHHFRILGEBQYNNRYOQYGAEBECVLOMGVLCPRPGCAGLLPEPDQRKTYCEGANGLCG 360
Db      301 LHHFRILGEBQYNNRYOQYGAEBECVLOMGVLCPRPGCAGLLPEPDQRKTYCEGANGLCG 360
Qy      361 GFAFCRECKEAYHEGSCSAVFEASGTTTOAYRVDERAAEQARWEAASKETIKKTYPCPR 420
Db      361 GFAFCRECKEAYHEGSCSAVFEASGTTTOAYRVDERAAEQARWEAASKETIKKTYPCPR 420
Qy      421 CHVPEKNGGCMHMKCPQPCRLEWCNCGCEMNRYCMGDHMFV 465
Db      421 CHVPEKNGGCMHMKCPQPCRLEWCNCGCEMNRYCMGDHMFV 465
```

```
RESULT 4
US-10-473-226-4
; Sequence 4, Application US/10473226
; Publication No. US20040198650A1
; GENERAL INFORMATION:
; APPLICANT: NcGene A/S
; TITLE OF INVENTION: Means for inhibiting proteolytical processing of Parkin
; FILE REFERENCE: 506-204-WO
; CURRENT APPLICATION NUMBER: US/10/473,226
; PRIOR FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: DK PA 2001 00525
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/281,286
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..(1)
; OTHER INFORMATION: Parkin with a D126E mutation
US-10-473-226-4

Query Match      99.5%; Score 2584; DB 16; Length 465;
Best Local Similarity 99.6%; Pred. No. 2.5e-219;
Matches 463; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIVFRFNSSHGFPVEVSDSTSI FOLKEVVARQGVADQLRVIFAGKELRNDMTVQNC 60
DB 1 MIVFRFNSSHGFPVEVSDSTSI FOLKEVVARQGVADQLRVIFAGKELRNDMTVQNC 60
QY 61 IDQGSIVHIVQRPWRKGEMNATGDDPRNAAAGCCEREPOSILTRVDLSSVLPQDSVGIA 120
DB 61 IDQGSIVHIVQRPWRKGEMNATGDDPRNAAAGCCEREPOSILTRVDLSSVLPQDSVGIA 120
QY 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
DB 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
QY 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
DB 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
QY 181 SCMDVLIIPNRMSGCQSPHCPTSAEFPKCGAHPSTSKETPVALLHATNSRNITCIT 240
DB 181 SCMDVLIIPNRMSGCQSPHCPTSAEFPKCGAHPSTSKETPVALLHATNSRNITCIT 240
QY 241 CTDVASPVLVFQCNRRHVICLDCFHLVCYTRLNDRQFVHDPOLGYSLFCVACCPNSLIKE 300
DB 241 CTDVASPVLVFQCNRRHVICLDCFHLVCYTRLNDRQFVHDPOLGYSLFCVACCPNSLIKE 300
QY 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
DB 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
QY 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
DB 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
QY 361 GFAFCRECKEAYHEGECSAVFASGTTTQAYRVDRAAQAQWAAASKETIKTKTPCPR 420
DB 361 GFAFCRECKEAYHEGECSAVFASGTTTQAYRVDRAAQAQWAAASKETIKTKTPCPR 420
QY 421 CHVPEKNGGCMHMKCPQPCRLBWCNCGCEMNRVCMGDHMFV 465
DB 421 CHVPEKNGGCMHMKCPQPCRLBWCNCGCEMNRVCMGDHMFV 465
```

```
RESULT 5
US-10-776-604-4
; Sequence 4, Application US/10776604
; Publication No. US20050003385A1
; GENERAL INFORMATION:
; APPLICANT: Shimizu, Nobuyoshi
; APPLICANT: Mizuno, Yoshiaki
; TITLE OF INVENTION: DNAs or Genes Participating in Parkinson's Disease
; FILE REFERENCE: 0652.2110001
; CURRENT APPLICATION NUMBER: US/10/776,604

RESULT 6
US-09-785-548-4
; Sequence 4, Application US/09785548
; Patent No. US2002015557A1
; GENERAL INFORMATION:
; APPLICANT: AVENTIS PHARMACEUTICALS, INC.
; TITLE OF INVENTION: COMPOSITIONS THAT CAN BE USED FOR REGULATING THE ACTIVITY OF PARKI
; FILE REFERENCE: ST00005
; CURRENT APPLICATION NUMBER: US/09/785,548
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-785-548-4

Query Match      99.4%; Score 870; DB 9; Length 156;
Best Local Similarity 99.4%; Pred. No. 1.3e-68;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIVFRFNSSHGFPVEVSDSTSI FOLKEVVARQGVADQLRVIFAGKELRNDMTVQNC 60
DB 1 MIVFRFNSSHGFPVEVSDSTSI FOLKEVVARQGVADQLRVIFAGKELRNDMTVQNC 60
QY 61 IDQGSIVHIVQRPWRKGEMNATGDDPRNAAAGCCEREPOSILTRVDLSSVLPQDSVGIA 120
DB 61 IDQGSIVHIVQRPWRKGEMNATGDDPRNAAAGCCEREPOSILTRVDLSSVLPQDSVGIA 120
QY 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
DB 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
QY 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
DB 121 VILHTDSRKDPPASPGASRSTYNSFYVCKGPCQVQPGKLRVOCSTCRQATLTLTGCP 180
QY 181 SCMDVLIIPNRMSGCQSPHCPTSAEFPKCGAHPSTSKETPVALLHATNSRNITCIT 240
DB 181 SCMDVLIIPNRMSGCQSPHCPTSAEFPKCGAHPSTSKETPVALLHATNSRNITCIT 240
QY 241 CTDVASPVLVFQCNRRHVICLDCFHLVCYTRLNDRQFVHDPOLGYSLFCVACCPNSLIKE 300
DB 241 CTDVASPVLVFQCNRRHVICLDCFHLVCYTRLNDRQFVHDPOLGYSLFCVACCPNSLIKE 300
QY 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
DB 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPCGAGLLPEPQRRVYTCGGNGLGC 360
QY 361 GFAFCRECKEAYHEGECSAVFASGTTTQAYRVDRAAQAQWAAASKETIKTKTPCPR 420
DB 361 GFAFCRECKEAYHEGECSAVFASGTTTQAYRVDRAAQAQWAAASKETIKTKTPCPR 420
QY 421 CHVPEKNGGCMHMKCPQPCRLBWCNCGCEMNRVCMGDHMFV 465
DB 421 CHVPEKNGGCMHMKCPQPCRLBWCNCGCEMNRVCMGDHMFV 465
```

QY 135 GSPAGRSIYNSFYVYCKGPCORVQPGKLRVOCSTCRQATLTITQSPSCMDVLLIPNMSG 194  
 DB 1 GSPAGRSIYNSFYVYCKGPCORVQPGKLRVOCSTCRQATLTITQSPSCMDVLLIPNMSG 60  
 QY 195 ECOSPHRCPTSAEFPFKCGAHPSTSDKETPVALHLIATNSRNITCTCTDVRSPLVFOCN 254  
 DB 61 ECOSPHRCPTSAEFPFKCGAHPSTSDKETPVALHLIATNSRNITCTCTDVRSPLVFOCN 120  
 QY 255 SRHVICLDCFHLVCTVRLNDROFVHDPOLGYSLECV 290  
 DB 121 SRHVICLDCFHLVCTVRLNDROFVHDPOLGYSLECV 156

RESULT 7  
 US-10-239-249-2  
 ; Sequence 2, Application US/10239249  
 ; Publication No. US20030177507A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HONER, MARTUS  
 ; APPLICANT: LINK, WOLFANG  
 ; APPLICANT: BAUMEISTER, RALF  
 ; TITLE OF INVENTION: NEMATODES AS MODEL ORGANISMS FOR INVESTIGATING  
 ; TITLE OF INVENTION: NEURODEGENERATIVE DISEASES AND, IN PARTICULAR, PARKINSON'S  
 ; TITLE OF INVENTION: DISEASE, USES AND METHODS FOR FINDING SUBSTANCES AND  
 ; TITLE OF INVENTION: GENES WHICH CAN BE USED IN TREATING SUCH DISEASES, AND  
 ; FILE REFERENCE: 02481.1804-00000  
 ; CURRENT APPLICATION NUMBER: US/10/239,249  
 ; CURRENT FILING DATE: 2002-09-20  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03214  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 100 14 109.9  
 ; PRIOR FILING DATE: 2000-03-22  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 386  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 US-10-239-249-2

Query Match  
 Best Local Similarity 21.0%; Score 544.5; DB 14; Length 386;  
 Matches 129; Conservative 59; Mismatches 166; Indels 93; Gaps 13;  
 QY 22 SIFQKEVVAARQGVADOLRVIPAGKELRNMTVONCULDQSIYHIVQPRKRGQEN 81  
 DB 29 NIEBLTKDVEKLTIRIPDELEVFCGKLSKSTIMDLTPATQIMLRPFNSHNENG 88  
 QY 82 ATGDDPRNNAAGGERPQSLTRVDLSSSVLPEDSVGLAVIIHTDSRKDPPAGSPAGS 141  
 DB 89 AT-----TAKITTD-----S 99  
 QY 142 IYNSFYVYCKGPCORVQPGKLRVOCSTCRQATLTITQSPSCMDVLLIPNMSGECOSPHC 201  
 DB 100 ILGSFYVYCKGN-CDDVVRGKLRYCCKGKSTVLYKSEFQNNSDVLSKRLPAVCEBCT 158  
 QY 202 PGTSAEFPFKCGAHPSTSDKETPVALHLIATNSRNITCTCTDVRSPLVFOCN SRHVICL 261  
 DB 159 PGLFAFPFKRC-----LACNDPAALTHVGNWQMECCVCGKKEVLPDLGN--HTTCQ 212  
 QY 262 DCHHLVCTVRLNDROFVHDPOLGYSLEPC-VAGCPNSLKEHLHRRILIGEOYRNYOQYA 320  
 DB 213 FCFRDLVLSQLRFRFVQNPRIHGTIFCPYRQC-NRVADVHNNHNGQTSYSYORKAT 271  
 QY 321 EECV-LQMGVLCPRPGGAGILPEP--DQRKVTCEGNGVLCGFAFCRCCKEAYHGE 377  
 DB 272 ERLIADVCKGVTCFVNSCGSGFWEPEYDDGRSQP-----DGFSPCRKC-----FERNC 322  
 QY 378 SAVFEASGTTTQAVYVDERAAEQARWEAASKETIKKTKRCPCGHVPEKNGSCGMKCP 437  
 DB 323 VCOSHDLTTRT-----TIDATIRCPKCHVATERNNGCAHIHC- 360

QY 438 QPQRLIEWCWNCGENNRVCMGDHFPD 464  
 DB 361 -TSCGMDWCFKCTEMKEECQMDHFN 386

RESULT 8  
 US-10-313-203-10  
 ; Sequence 10, Application US/10313203  
 ; Publication No. US20040029134A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gu, Wei  
 ; APPLICANT: Nikolaev, Anatoly  
 ; TITLE OF INVENTION: P53-ASSOCIATED PARKIN-LIKE CYTOPLASMIC PROTEIN, AND RELATED COMPO  
 ; TITLE OF INVENTION: AND METHODS  
 ; FILE REFERENCE: 68106  
 ; CURRENT APPLICATION NUMBER: US/10/313,203  
 ; CURRENT FILING DATE: 2002-12-06  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 10  
 ; LENGTH: 64  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-313-203-10

Query Match  
 Best Local Similarity 14.4%; Score 373; DB 15; Length 64;  
 Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 314 RYQOYGAEECYLQMGVLCPRPGCGAGLLPEPDQRKVTCEGNGVLCGFAFCRCCKEAYH 373  
 DB 1 RYQOYGAEECYLQMGVLCPRPGCGAGLLPEPDQRKVTCEGNGVLCGFAFCRCCKEAYH 60

QY 374 EGEC 377  
 DB 61 EGEC 64

RESULT 9  
 US-10-313-203-17  
 ; Sequence 17, Application US/10313203  
 ; Publication No. US20040029134A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gu, Wei  
 ; APPLICANT: Nikolaev, Anatoly  
 ; TITLE OF INVENTION: P53-ASSOCIATED PARKIN-LIKE CYTOPLASMIC PROTEIN, AND RELATED COMPO  
 ; TITLE OF INVENTION: AND METHODS  
 ; FILE REFERENCE: 68106  
 ; CURRENT APPLICATION NUMBER: US/10/313,203  
 ; CURRENT FILING DATE: 2002-12-06  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 17  
 ; LENGTH: 56  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-313-203-17

Query Match  
 Best Local Similarity 12.4%; Score 323; DB 15; Length 56;  
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 CITCTDVRSPLVFOCN SRHVICLDCFHLVCTVRLNDROFVHDPOLGYSLECVAGC 293  
 DB 1 CITCTDVRSPLVFOCN SRHVICLDCFHLVCTVRLNDROFVHDPOLGYSLECVAGC 56

RESULT 10  
 US-09-864-761-36750  
 ; Sequence 36750, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:34:50 / Search time 19.3575 Seconds  
(without alignments)  
2311.294 Million cell updates/sec

Title: US-10-622-817-9  
Perfect score: 2596  
Sequence: 1 MIVFRRFNSHGFPVEVDSD.....CMNCGCEWNRVCMGDHFDV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	490	18.9	357	2	T23460 hypothetical prote
2	220	8.5	543	2	H84724 probable ARI-like
3	219.5	8.5	1753	2	T00350 hypothetical prote
4	205	7.9	437	2	T16477 hypothetical prote
5	205	7.9	491	2	F87793 protein C27A12.6
6	198.5	7.6	514	2	T02366 hypothetical prote
7	196.5	7.6	514	2	G84724 probable ARI-like
8	195	7.5	320	2	T45909 hypothetical prote
9	193	7.4	542	2	H84725 similar to Atrialde
10	190.5	7.3	436	2	H87793 protein C27A12.8
11	188	7.2	1048	2	T13653 hypothetical prote
12	186.5	7.2	324	2	T16983 hypothetical prote
13	182.5	7.0	551	2	S38086 hypothetical prote
14	180.5	6.9	594	2	T04783 protein C27A12.7
15	178.5	6.9	497	2	G87793 protein C27A12.7
16	178.5	6.9	497	2	B86448 hypothetical prote
17	174.5	6.7	408	2	T47498 hypothetical prote
18	174	6.7	348	2	T47498 hypothetical prote
19	173	6.7	565	2	F84721 probable RING zinc
20	162.5	6.3	498	2	TC5983 protein kinase C-1
21	159	6.1	518	2	E84536 hypothetical prote
22	158.5	6.1	451	2	T16481 hypothetical prote
23	157	6.0	488	2	T29562 hypothetical prote
24	155	6.0	532	2	T04748 hypothetical prote
25	152.5	5.9	816	2	T25555 hypothetical prote
26	149	5.7	464	2	S48329 probable membrane
27	147	5.7	1209	2	T52523 ubiquitin / riboso
28	143.5	5.5	155	2	JH0227 ubiquitin / riboso
29	142.5	5.5	155	2	JH0226 ubiquitin / riboso

30	141.5	5.5	468	2	A84601 Mutator-like trans
31	136	5.2	222	2	T47487 hypothetical prote
32	134.5	5.2	229	1	UCUNI ubiquitin precuro
33	134	5.2	373	2	B84647 hypothetical prote
34	134	5.2	638	2	T02611 hypothetical prote
35	133	5.1	397	2	H84578 probable RING zinc
36	132.5	5.1	869	1	JC4858 VLDL receptor prec
37	132	5.1	80	2	T28305 ORF MSY144 probabl
38	132	5.1	150	2	T30390 probable ubiquitin
39	132	5.1	154	2	S55242 polyubiquitin 2 -
40	132	5.1	156	2	C86439 protein T19E23.13
41	131	5.0	534	2	S34285 polyubiquitin - to
42	130.5	5.0	305	1	S29853 polyubiquitin 4 -
43	130.5	5.0	356	1	UCUTRC polyubiquitin / r1
44	130	5.0	155	2	S40240 ubiquitin/ribosoma
45	130	5.0	156	2	T52335 ubiquitin extensio

ALIGNMENTS

RESULT 1	T23460	hypothetical protein K083.7 - Caenorhabditis elegans
C:Species:	Caenorhabditis elegans	
C:Date:	15-Oct-1999	#sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession:	T23460	
R:McMurray, A.		
	submitted to the EMBL Data Library, November 1996	
A:Reference number:	Z19743	
A:Accession:	T23460	
A:Status:	preliminary; translated from GB/EMBL/DBJ	
A:Molecule type:	DNA	
A:Residues:	1-357 <WIL>	
A:Cross-References:	UNIPROT:O9XUS3, EMBL:Z81568, PIDN:CAB04599.1, GSPDB:GN00021, CESP:K01	
A:Experimental source:	clone K08E3	
C:Genetics:		
A:Gene:	CESP:K08E3.7	
A:Map position:	3	
A:introns:	23/3; 72/2; 218/1; 286/2; 311/2	
Query Match	18.9%; Score 490; DB 2; Length 357;	
Best Local Similarity	26.7%; Pred. No. 1.8e-29;	
Matches	119; Conservative 53; Mismatches 154; Indels 120; Gaps 12;	
QY	22	STFOLKEVAVKQGVPAQOLRYIFPAGKELRNDWTQNCDDLDOQSIVHTVQREWRKQEQEM 81
DB	29	NIEDLTKVKEKLTETPSDELVEVFCGKKLSKSTIRDLSTPATQIMLRPFNSHNENG 88
QY	82	ATGSDPPRNAAGCGRREPQSLTRVDLSSVLPFGDSVGLAVILHTPSRKDSPAGSPAGRS 141
DB	89	AT-----TAKTTTS-----S 99
QY	142	IYNSFYVYCKGPGQROVQPKLRVQSCGTCQATLTLTGSPSCWDVLIIPRMGEGCQSPHC 201
DB	100	ILGSFYVWCKN-CDVVKRGLKRLAVYQCKCSSTSIVLKSEFQNMVDVLSKSRITPAVCECCT 158
QY	202	PQTSAEFFPKCGAHPSTDKETPVALLIATNSRNITCTCTVRSBVLVFOGNSRNVICL 261
DB	159	PGLFAEFKFKC---LACNDPAALTHVGNQMTECCVCDKEKVIPLDGCN--HITQ 212
QY	262	DCFHLVCTRLNRDQFVHDPOLGYSLLPCVAGPNSLILKHHFRILIGEEVRYOOGYAE 321
DB	213	FCFRV-----RVDDVHHFHMGQSYSEYQKATE 243
QY	332	ECV-LQMGGLCPRGCGAGLPRP--DQRYTCGNGNLGCGAFRCBCKEAYHEGCS 378
DB	244	RLIAYDDKGVTCPPNNSCGGSPFWEYDDGRGQC---DCFSPCRKC---FERNCV 294
QY	379	AVFEASGTTTQAYRVDERAAEQARWEASKEITTKTTPCPCHYPVENGGCMKCKQ 438
DB	295	COSEDDLTRT-----TIDATTRCPCHVATERNGGCAIHIC-- 331

QY 439 POCRLMCMNCCGEMNRVCMGDHWD 464  
DB 332 TSCGMDWCFKCKTEKKECCQMDHMFN 357

RESULT 2

probable ARI-like RING zinc finger protein (imported) - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: H847724  
R/Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentol, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.;  
Nuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: H847724  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-543 <STO>  
A/Cross-references: UNIPROT:Q9SKC3; GB:AE002093; NID:94887759; PIDN:BA032295.1; GSPDB:GN  
C/Genetics:  
A/Map position: 2

Query Match  
Best Local Similarity 25.9%; Score 220; DB 2; Length 543;  
Matches 60; Conservative 33; Mismatches 83; Indels 56; Gaps 11;

QY 235 NITCITCTDVRSPVVFQCNHRVCLDCHLYCTRLND-----ROFVHPQGLYSLPC 289  
DB 125 NIQCGICFESYREIARVSCHPYCKTCMAGYITTKIEDPGCLRYKCEPS-----C 178  
QY 290 VAGCNSLIKEIHFRIIGEEQYNNY--QOYGAECVLOMGVLCPPRGAGGLT---PE 344  
DB 179 SAAVKGMIEDVETKVV--NEKYSRYILRSYVEDGKKIKW-----CPSPGCGYAVVERGGSS 232  
QY 345 PDQKVTCEGNGAGCGAFCECKEAVHEG--ESAVFEASGTTQAYRDERAERAEQARW 403  
DB 233 SSSYVSLT-----CSTRFCMNCSEDNHSPVDDTV-----SKW 266  
QY 404 -----EAAKERTTKTKTPCRCHVPEKNGGCMHKCPQPCRLEMCNMC 449  
DB 267 IFKNQDESENKMWLANSKPCBECKPRIEKNDGNMTGAP--CGHEFCMTC 317

RESULT 3

hypothetical protein KIAA0708 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C/Accession: T00350  
R/Linikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.  
DNA Res. 5, 169-176, 1998  
A/Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A/Reference number: Z14142; MUID:98403880; PMID:9734811  
A/Accession: T00350  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1753 <ISH>  
A/Cross-references: UNIPROT:Q8IWT3; EMBL:AB014608; NID:93327229; PIDN:BA031683.1; PID:93  
A/Experimental source: brain  
C/Genetics:  
A/Note: KIAA0708

Query Match  
Best Local Similarity 8.5%; Score 219.5; DB 2; Length 1753;  
Matches 125; Conservative 57; Mismatches 181; Indels 221; Gaps 32;

QY 1 MIVPFRNNSSGFPEV---DSPTS-----IFOLK 27  
DB 1012 MWLTKENQTEVSEVETLLKSDLSPELLLQALVPLTSGNGPLTLHEGQDFRIGVLRLLH 1071

QY 28 EVAAKQO-----VPADQLRVIFAKETLRNDMTV-----QNCDDDOOSI 66  
DB 1072 EPGPORSGEALMLIPPOAYLNVKEKGRITLQKRNLSGLVRIILKAHGEKGLHTDQ--L 1129

RESULT 4

67 VHIYORPMRGOMNATGDDPRNNAAGCEREPOLITRVDSVLPFGDSVGLAVILHTD 126  
DB 1130 VCVLTEMQGNRPPTGLG--HTVAGV-----ACTSTDVLSCTL--HLGGQVY---- 1175  
QY 127 SRKDSPP-----AGSPAGRSIYNSFYVCKPQCRVOPGRLRYOCSCTCRQATLTITGCP 180  
DB 1176 KRDRDRPOLIMVYAPFPWPGCGADVPCGSGASTSPSPRAV-----ATLASQLP 1228  
QY 181 SCWDVLLINRMBSGEGQSPHCSTAEFFPK--CGAHPSTDKETPAVHLIATNS---- 233  
DB 1229 -----AGRTMSFQ--EVEGIMQYVRQVETLNLBPDAVQHLAHSHMGAE 1272  
QY 234 -----RNITCITCTDVRSPV-----LVQCNHRVHT 259  
DB 1273 QLLQSYSDPEPDLILAAGLCVHQAAVAVVRPDHPCV--SPUGCDDDLPSLC--CMHYC 1328  
QY 260 CLDCHLYCTRLNDROFVHPDQGLYSLPC--VAGCP-----NSLIKEIHFRIIGEEQYNN 314  
DB 1329 CASKNNEYLTTRI-----EONLVINCTCPICADCPAQPTGAFIRAL--VSPVVISK 1377  
QY 315 YQO-----YGAEECVLOMGVLCPRP--GCCAGLLPBDQKVTCEGNGLG-----CGFA 363  
DB 1378 YEKALLRGY--VESC--SNLTWCTNPGCD-----RLICRQGLCGYTCCKGWA 1423  
QY 364 FCRECK--EAVHBECSAVFEASGTTQAYRDERAERAEQARW-----EASKE 409  
DB 1424 SCFNCSPFEAHYVASC-----GHMSQWDDOGGYDGKSVBQSHK 1463

QY 410 TIKKTRPCRCCHVPEKNGGCMHKCPQPCRLEMCNMCCEW 453  
DB 1464 LAKLISRCPCSCQAPIRKHEGLHMTG--AKCNHGCWRLKSW 1505

RESULT 4

hypothetical protein F56D2.5 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T16477  
R/Du, Z.  
submitted to the EMBL Data Library, August 1994  
A/Description: The sequence of C. elegans cosmid F56D2.  
A/Reference number: Z18519  
A/Accession: T16477  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-437 <DUZ>  
A/Cross-references: UNIPROT:Q20871; EMBL:U13644; NID:9532100; PID:91945502; PIDN:AA05268;  
C/Genetics:  
A/Experimental source: strain Bristol N2; clone F56D2  
A/Map position: 3  
A/Introns: 41/3; 134/3; 185/3; 223/3; 254/2; 287/2; 414/3

Query Match  
Best Local Similarity 7.9%; Score 205; DB 2; Length 437;  
Matches 104; Conservative 66; Mismatches 169; Indels 176; Gaps 28;

QY 18 DSDTISIFOLKEVAVKRGVADQLRVIFAKEL--RNDVTVONCDIDQOSIVHIQRPWR 75  
DB 3 DRDQIYEL-----EALSVLEKRLKANSKSDMSDKNAET--QGIIEV----- 42  
QY 76 KQGMNATGDDPRNNAAGCEREPOLITRVDSVLPFGDSVGLAVILHTDSRKXSPAG 135  
DB 43 -----GFDNLVYDPTVITEGTSDSGDFHLPLDILP-----PIRLKFLHPNDYPTVS 88  
QY 136 SPAGRSIYNSFYVCKPQCRVOPGRLRYOCSCTCRQATLTITGQSCWDVLLINRMBSGE 195



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2005, 13:33:04 ; Search time 88.371 Seconds  
(without alignments)  
2694.512 Million cell updates/sec

Title: US-10-622-817-9  
Perfect score: 2596  
Sequence: 1 MIVFRRFNSSHGPFVVDSD.....CWNCGCEMRNVCMDHMFDPV 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2596	100.0	465 1 PRKN_HUMAN	060260 homo sapien
2	2218	85.4	465 1 PRKN_RAT	091666 rattus norv
3	2170.5	83.6	464 1 PRKN_MOUSE	094966 mus musculu
4	1115	43.0	203 2 OBN142	081142 homo sapien
5	1060	40.8	194 2 Q95M07	091607 bos taurus
6	1058.5	40.8	2 Q95M14	095114 drosophila
7	1058.5	40.8	482 2 Q7KTX7	071637 drosophila
8	1049	40.4	8 070591	074591 anopheles g
9	1024	39.4	177 2 OBN141	081141 homo sapien
10	957	36.9	177 2 OBN142	081142 rattus norv
11	544.5	21.0	386 2 Q9XUS3	094963 caenorhabd
12	543.5	20.9	162 2 Q86LE7	086167 drosophila
13	485	18.7	111 2 OBN145	081145 rattus norv
14	420.5	16.2	120 2 O7TPE8	071638 mus musculu
15	382	14.7	78 2 OBN145	081145 mus musculu
16	284	10.9	50 2 O6Q216	061216 homo sapien
17	256	9.9	520 2 Q84R80	084186 dictyostell
18	250.5	9.6	562 2 Q84R80	084186 aradidopsis
19	249.5	9.6	562 2 Q84R80	084186 xenopus lae
20	246	9.5	525 2 O653S8	065358 cryza sativ
21	240.5	9.3	503 1 AR11_DROME	094981 drosophila
22	240	9.2	492 2 Q71355	071355 brachydanto
23	240	9.2	511 2 Q9VU94	094994 anopheles g
24	238.5	9.1	2 Q70G07	070G07 anopheles g
25	236.5	9.1	445 2 O8CFJ4	08CFJ4 mus musculu
26	236.5	9.1	533 2 O6NM85	06NM85 brachydanto
27	236.5	9.1	555 1 AR11_MOUSE	094115 mus musculu
28	236.5	9.1	557 1 AR11_HUMAN	094115 homo sapien
29	234.5	9.0	527 2 Q6PFJ9	06PFJ9 brachydanto
30	232.5	8.9	492 1 AR12_MOUSE	094116 mus musculu
31	231	8.9	491 2 Q6GL03	06GL03 xenopus tro

32	231	8.9	493 1 AR12_HUMAN	095376 homo sapien
33	231	8.9	607 2 Q6Z1V0	06Z1V0 cryza sativ
34	227.5	8.8	589 2 Q7XRI6	07XRI6 cryza sativ
35	225.5	8.7	501 2 Q6CSC0	06CSC0 Kluveromyc
36	225.5	8.7	2489 2 Q68CP2	068CP2 homo sapien
37	224	8.6	551 2 Q84R11	084R11 aradidopsis
38	224	8.6	552 2 Q81829	081829 aradidopsis
39	222	8.6	474 1 RN14_HUMAN	094968 homo sapien
40	222	8.6	474 2 Q61BVO	061BVO homo sapien
41	222	8.6	492 2 Q6R2U9	06R2U9 musa acum
42	220	8.5	40 2 Q6S8G7	06S8G7 homo sapien
43	220	8.5	543 2 Q9SKC3	09SKC3 aradidopsis
44	219.5	8.5	2517 1 PARC_HUMAN	081WC3 homo sapien
45	219.5	8.5	2517 2 Q68D92	068D92 homo sapien

## ALIGNMENTS

RESULT 1  
PRKN\_HUMAN STANDARD; PRT; 465 AA.  
ID PRKN\_HUMAN  
AC 060260, OBN143, OBN144, Q9M07;  
DT 25-OCT-2004 (Rel. 45, Created)  
DT 25-OCT-2004 (Rel. 45, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE parkin (EC 6.3.2.-) (Ubiquitin E3 ligase PRKN) (Parkinson juvenile disease protein 2) (Parkinson disease protein 2).  
GN Name=PRKN; Synonyms=PRKN;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INVOLVEMENT IN JUVENILE PARKINSON'S DISEASE.  
RC TISSUE=Fetal brain, and Skeletal muscle;  
RX MEDLINE=98219084; PubMed=9560156; DOI=10.1038/33416;  
RA Kitada T., Asakawa S., Hattori N., Matsunaga H., Yamamura Y., Minoshima S., Yokochi M., Mizuno Y., Shimizu N.,  
RT "Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.";  
RL Nature 392:605-608(1998).  
RN  
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
RA D'Agata V., Scapagnini G., Cavallaro S.,  
RT "Functional and molecular diversity of parkin.";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN  
RP SEQUENCE FROM N.A. (ISOFORM 5).  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,  
RA Altshuler S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stempleton M., Soares W.B., Bonaldo M.F., Casavani T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Locantore N.A., Peters G.J., Aberson R.D., Mallory S.J.,  
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Morley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield V.S.N., Krzywicki M.I., Skalek U., Smalls D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Motta M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN

RP FUNCTION IN UBIQUITINATION.  
 RX PubMed=10973942; DOI=10.1074/jbc.C000447200;  
 RA Imai Y., Soda M., Takahashi R.;  
 RT "Parkin suppresses unfolded protein stress-induced cell death through  
 its E3 ubiquitin-protein ligase activity.";  
 RL Nat. Chem. 275:35661-35664(2000).  
 RN [5]

RP FUNCTION, AND CHARACTERIZATION OF VARIANTS PD PRO-42 AND ARG-240.  
 RX PubMed=10888878; DOI=10.1038/77060;  
 RA Shimura H., Hattori N., Kubo S.-I., Mizuno Y., Asakawa S.,  
 RA Minoshima S., Shimizu N., Iwai K., Chiba T., Tanaka K., Suzuki T.;  
 RT "Familial Parkinson disease gene product, parkin, is a ubiquitin-  
 protein ligase.";  
 RL Nat. Genet. 25:302-305(2000).  
 RN [6]

RP FUNCTION, AND INVOLVEMENT IN CANCER.  
 RX PubMed=12719539; DOI=10.1073/pnas.0931262100;  
 RA Cesari R., Martin E.S., Calin G.A., Pentimalli F., Bichi R.,  
 RA McAdams H., Trapasso F., Drusco A., Shimizu M., Masciullo V.,  
 RA D'Andrilli G., Scambia G., Picchio M.C., Alder H., Godwin A.K.,  
 RT "Parkin, a gene implicated in autosomal recessive juvenile  
 parkinsonism, is a candidate tumor suppressor gene on chromosome 6q25-  
 q27.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5956-5961(2003).  
 RN [7]

RP FUNCTION, INTERACTIONS WITH FBXW7 AND CUL1, AND UBIQUITINATION OF  
 CYCLIN E.  
 RX PubMed=12628165;  
 RA Staropoli J.F., McDermott C., Martinat C., Schulman B., Demireva E.,  
 RA Abellovich A.;  
 RT "Parkin is a component of an SCF-like ubiquitin ligase complex and  
 RT protects postmitotic neurons from kainate excitotoxicity.";  
 RL Neuron 37:735-749(2003).  
 RN [8]

RP FUNCTION, UBIQUITINATION, AND S-NITROSYLATION.  
 RX PubMed=15105460; DOI=10.1126/science.1093891;  
 RA Chung K.K., Thomas B., Li X., Pletnikova O., Troncoso J.C., Marsh L.,  
 RA Dawson V.L., Dawson T.M.;  
 RT "S-nitrosylation of parkin regulates ubiquitination and compromises  
 RT parkin's protective function.";  
 RL Science 304:1328-1331(2004).  
 RN [9]

RP SUBCELLULAR LOCATION.  
 RX PubMed=10319893;  
 RA Shimura H., Hattori N., Kubo S.-I., Yoshikawa M., Kitada T.,  
 RA Matsumine H., Asakawa S., Minoshima S., Yamamura Y., Shimizu N.,  
 RA Mizuno Y.;  
 RT "Immunohistochemical and subcellular localization of Parkin protein:  
 RT absence of protein in autosomal recessive juvenile parkinsonism  
 RL patients.";  
 RL Ann. Neurol. 45:668-672(1999).  
 RN [10]

RP INTERACTIONS WITH UBE2L6 AND SEPT5, AND UBIQUITINATION OF SEPT5.  
 RX PubMed=11078524; DOI=10.1073/pnas.240347997;  
 RA Zhang Y., Gao J., Chung K.K., Huang H., Dawson V.L., Dawson T.M.;  
 RT "Parkin functions as an E2-dependent ubiquitin-protein ligase and  
 RT promotes the degradation of the synaptic vesicle-associated protein,  
 RT CDCrel-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13354-13359(2000).  
 RN [11]

RP UBIQUITINATION OF GPR37.  
 RX MEDLINE=21332586; PubMed=11439185; DOI=10.1016/S0092-8674(01)00407-X;  
 RA Imai Y., Soda M., Inoue H., Hattori N., Mizuno Y., Takahashi R.;  
 RT "An unfolded putative transmembrane polypeptide, which can lead to  
 RT endoplasmic reticulum stress, is a substrate of Parkin.";  
 RL Cell 105:891-902(2001).  
 RN [12]

RP INTERACTION, UBIQUITINATION OF SNCAIP, CHARACTERIZATION OF VARIANTS  
 RP ARG-240, CYG-256, TRP-275 AND ASN-415, AND MUTAGENESIS OF  
 RP CYG-337, CYG-421 AND CYG-431.  
 RX PubMed=11590439; DOI=10.1038/nm1001-1144;  
 RA Chung K.K., Zhang Y., Lim K.L., Tanaka Y., Huang H., Gao J.,  
 RA

RA Ross C.A., Dawson V.L., Dawson T.M.;  
 RT "Parkin ubiquitinates the alpha-synuclein-interacting protein,  
 RT synphilin-1: implications for Lewy-body formation in Parkinson  
 RT disease.";  
 RL Nat. Med. 7:1144-1150(2001).  
 RN [13]

RP INTERACTIONS WITH STUB1 AND HSP70, AND UBIQUITINATION OF STUB1.  
 RX PubMed=12150907;  
 RA Imai Y., Soda M., Harakeyama S., Akagi T., Hashikawa T.,  
 RA Nakayama K.-I., Takahashi R.;  
 RT "CHIP is associated with Parkin, a gene responsible for familial  
 RT Parkinson's disease, and enhances its ubiquitin ligase activity.";  
 RL Mol. Cell 10:55-67(2002).  
 RN [14]

RP INTERACTION WITH SYT11, CHARACTERIZATION OF VARIANT PD GLY-289, AND  
 RP MUTAGENESIS OF CYG-418  
 RX PubMed=12925569; DOI=10.1093/hmg/ddg269;  
 RA Huynh D.P., Scoles D.R., Nguyen D., Pulst S.M.;  
 RT "The autosomal recessive juvenile Parkinson disease gene product,  
 RT parkin, interacts with and ubiquitinates synaptotagmin XI.";  
 RL Hum. Mol. Genet. 12:2587-2597(2003).  
 RN [15]

RP UBIQUITINATION OF AN O-LINKED GLYCOSYLATED ISOFORM OF SNCAIP,  
 RP SUBCELLULAR LOCATION, AND CHARACTERIZATION OF VARIANTS PD PRO-42 AND  
 RP ARG-240.  
 RX PubMed=11431533; DOI=10.1126/science.1060627;  
 RA Shimura H., Schlossmacher M.G., Hattori N., Froesch M.P.,  
 RA Trockenbacher A., Schneider R., Mizuno Y., Kozik K.S., Selkoe D.J.;  
 RT "Ubiquitination of a new form of alpha-synuclein by parkin from human  
 RT brain: implications for Parkinson's disease.";  
 RL Science 293:63-69(2001).  
 RN [16]

RP INVOLVEMENT IN CANCER.  
 RX PubMed=14614460; DOI=10.1038/sj.onc.1207072;  
 RA Denison S.R., Wang F., Becker N.A., Schuele B., Kock N.,  
 RA Phillips L.A., Klein C., Smith D.I.;  
 RT "Alterations in the common fragile site gene Parkin in ovarian and  
 RT other cancers.";  
 RL Oncogene 22:8370-8378(2003).  
 RN [17]

RP REVIEW.  
 RX PubMed=15229644; DOI=10.1038/sj.embor.7400188;  
 RA Kahle P.J., Haass C.;  
 RT "How does parkin ligase ubiquitin to Parkinson's disease?";  
 RL EMBO Rep. 5:681-685(2004).  
 RN [18]

RP STRUCTURE BY NMR OF 1-76, AND INTERACTION WITH PSMD4.  
 RX PubMed=12634850; DOI=10.1038/sj.embor.embor764;  
 RA Sakata E., Yamaguchi Y., Kurimoto E., Kikuchi J., Yokoyama S.,  
 RA Yamada S., Kawahara H., Yokosawa H., Hattori N., Mizuno Y., Tanaka K.,  
 RA Kato K.;  
 RT "Parkin binds the Rpn10 subunit of 26S proteasomes through its  
 RT ubiquitin-like domain.";  
 RL EMBO Rep. 4:301-306(2003).  
 RN [19]

RP REVIEW ON VARIANTS.  
 RX PubMed=14976155; DOI=10.1093/hmg/ddh089;  
 RA Mata I.F., Lockhart P.J., Farrer M.J.;  
 RT "Parkin genetics: one model for Parkinson's disease.";  
 RL Hum. Mol. Genet. 13:R127-R133(2004).  
 RN [20]

RP VARIANT ARG-240.  
 RX PubMed=9731209;  
 RA Hattori N., Matsumine H., Asakawa S., Kitada T., Yoshino H.,  
 RA Elibol B., Brookes A.J., Yamamura Y., Kobayashi T., Wang M.,  
 RA Yoritaka A., Minoshima S., Shimizu N., Mizuno Y.;  
 RT "Point mutations (Thr240Arg and Glu31Stop) in the Parkin gene.";  
 RL Biochem. Biophys. Res. Commun. 249:754-758(1998).  
 RN [21]

RP ERRATUM.  
 RA Hattori N., Matsumine H., Asakawa S., Kitada T., Yoshino H.,  
 RA Elibol B., Brookes A.J., Yamamura Y., Kobayashi T., Wang M.,  
 RA Yoritaka A., Minoshima S., Shimizu N., Mizuno Y.;

BL Biochem. Biophys. Res. Commun. 251:666-666(1998).  
 RN [122]  
 Query Match 100.0%; Score 2596; DB 1; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-195;  
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIVFRFNSHGFPVEVSDTSTIPOLKEVAVKQGVAPDLVITPAGKLRNDMTVQND 60  
 DB 1 MIVFRFNSHGFPVEVSDTSTIPOLKEVAVKQGVAPDLVITPAGKLRNDMTVQND 60  
 QY 61 LDQGGIVHIVQRPWKRGQEMATGDDPRNAAAGCEREPOSITRYDLSSTVLPDGVGLA 120  
 DB 61 LDQGGIVHIVQRPWKRGQEMATGDDPRNAAAGCEREPOSITRYDLSSTVLPDGVGLA 120  
 QY 121 VILHDSKRDSPAPASPRASITNSFYVCKPQCRVQPGKLRVQSTCRQATLTLTQGP 180  
 DB 121 VILHDSKRDSPAPASPRASITNSFYVCKPQCRVQPGKLRVQSTCRQATLTLTQGP 180  
 QY 181 SCMDVLLPNRMSGSCSPHCPTSAEPFKCAHPTSKETPVALLHLATNSRNTCTT 240  
 DB 181 SCMDVLLPNRMSGSCSPHCPTSAEPFKCAHPTSKETPVALLHLATNSRNTCTT 240  
 QY 241 CTDVRSPLYVPCQNSRHVICTDCEFLHYCVTRLNDRQFVHPDQGYSLPCVACGPNLSLKE 300  
 DB 241 CTDVRSPLYVPCQNSRHVICTDCEFLHYCVTRLNDRQFVHPDQGYSLPCVACGPNLSLKE 300  
 QY 301 LHHFPIILBEOQYRVOQGAEECVLOMGVLCPRPGCGAGLLPEPDQRYTEGGNGLGC 360  
 DB 301 LHHFPIILBEOQYRVOQGAEECVLOMGVLCPRPGCGAGLLPEPDQRYTEGGNGLGC 360  
 QY 361 GPAPRECKEAVHEBECSAVFASGTTTQAYRVDEBAAQARWEAASKETIKTKPCPR 420  
 DB 361 GPAPRECKEAVHEBECSAVFASGTTTQAYRVDEBAAQARWEAASKETIKTKPCPR 420  
 QY 421 CHAVPEKNGGCMHMKCPOPOCRLEMCNMGCEMNRCVMDHNPV 465  
 DB 421 CHAVPEKNGGCMHMKCPOPOCRLEMCNMGCEMNRCVMDHNPV 465  
 RESULT 2  
 PRKN RAT STANDARD; PRT: 465 AA.  
 ID \_PRKN RAT STRAND; PRT: 465 AA.  
 AC 25-OCT-2004 (Rel. 45, Created)  
 DT 25-OCT-2004 (Rel. 45, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Parkin (EC 6.3.2.-) (Ubiquitin E3 ligase PRKN).  
 GN Name=Park2; Synonyms=Prkn;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=20153712; PubMed=10686358; DOI=10.1016/S0169-328X(99)00286-7;  
 RA D'Agata V., Zhao W., Cavallaro S.;  
 RT "Cloning and distribution of the rat parkin mRNA."  
 RL Brain Res. Mol. Brain Res. 75:345-349(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RC TISSUE=hypothalamus;  
 RX MEDLINE=2019698; PubMed=10737637;  
 RA Gu W.-D., Abbas N., Lagunes M.Z., Parent A., Pradier L., Bohne G.A.,  
 RA Agid Y., Hirsch E.C., Ragsman-Vozari R., Brice A.;  
 RT "Cloning of rat parkin cDNA and distribution of parkin in rat brain."  
 RL J. Neurochem. 74:1773-1776(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=Sprague-Dawley;  
 RA Hattori N., Wang M., Mizuno Y.;  
 RT "The expression of parkin mRNA in developing, adult and ageing rat

RT CNS.";  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RA Soda M., Imai Y., Takahashi R.;  
 RT "Molecular cloning of rat parkin gene."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Functions within a multiprotein E3 ubiquitin ligase  
 CC complex, catalyzing the covalent attachment of ubiquitin moieties  
 CC onto substrate proteins. These substrates include SYT11, C9orf1,  
 CC GPR37, STUB1, a 22 kDa O-linked glycosylated isoform of SNCAIP and  
 CC SEPT5. May play a more general role in the ubiquitin proteasomal  
 CC pathway by participating in the removal and/or deoxidation of  
 CC abnormally folded or damaged protein. Loss of this ubiquitin  
 CC ligase activity appears to be the mechanism underlying  
 CC pathogenesis of AR-OP. May protect neurons against alpha synuclein  
 CC toxicity, proteasomal dysfunction, GPR37 accumulation, and  
 CC kainate-induced excitotoxicity. May play a role in controlling  
 CC neurotransmitter trafficking at the presynaptic terminal and in  
 CC calcium-dependent exocytosis. Regulates cyclin B during neuronal  
 CC apoptosis. May represent a tumor suppressor gene (By similarity).  
 CC -1- PATHWAY: Ubiquitin conjugation; third step.  
 CC -1- SUBUNIT: Forms an E3 ubiquitin ligase complex with UBE2L3 or  
 CC UBE2L6. Part of a SCF-like complex, consisting of PARK2, CUL1 and  
 CC FBXW7. Interacts with SNCAIP. Binds to the C2A and C2B domains of  
 CC SYT11. Interacts and regulates the turnover of SEPT5. Part of a  
 CC complex, including STUB1, HSP70, and GPR37. The amount of STUB1 in  
 CC the complex increases during ER stress. STUB1 promotes the  
 CC dissociation of HSP70 from PARK2 and GPR37, thus facilitating  
 CC PARK2-mediated GPR37 ubiquitination. HSP70 transiently associates  
 CC with unfolded GPR37 and inhibits the E3 activity of PARK2,  
 CC whereas, STUB1 enhances the E3 activity of PARK2 through promotion  
 CC of dissociation of HSP70 from PARK2-GPR37 complexes. Interacts  
 CC with BSM24 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Expressed in the endoplasmic  
 CC reticulum, dendrites, some presynaptic terminals and in  
 CC postsynaptic densities (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Name=1;  
 CC IsoId=G9JK66-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=G9JK66-2; Sequence=VSP\_011722, VSP\_011723;  
 CC Name=3;  
 CC IsoId=G9JK66-3; Sequence=VSP\_011717;  
 CC Name=4;  
 CC IsoId=G9JK66-4; Sequence=VSP\_011718;  
 CC Name=5;  
 CC IsoId=G9JK66-5; Sequence=VSP\_011719;  
 CC Name=6;  
 CC IsoId=G9JK66-6; Sequence=VSP\_011717, VSP\_011720, VSP\_011721;  
 CC -1- TISSUE SPECIFICITY: Largely confined to neuronal elements,  
 CC including fibers and neuropil. Highly expressed at the forebrain  
 CC level, in pyramidal cells of layer V, in various cortical regions  
 CC and cerebellum. Expressed in the nucleus of diagonal band of  
 CC Broca, nucleus basalis, bed nucleus of the stria terminalis, and  
 CC olfactory tubercle. Moderate expression is seen in most neurons of  
 CC the subthalamic nucleus, heart, skeletal muscle and testis.  
 CC Moderate expression was found in frontal cortex, parietal cortex,  
 CC cerebellum, heart, skeletal muscle and testis.  
 CC -1- DOMAIN: The ubiquitin-like domain binds the PSD4 subunit of 26S  
 CC proteasomes (By similarity).  
 CC -1- PTM: Auto-ubiquitinates in an E2-dependent manner leading to its  
 CC own degradation (By similarity).  
 CC -1- PTM: S-nitrosylated (By similarity).  
 CC -1- SIMILARITY: Contains 2 IRR-type zinc fingers.  
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC -1- CAUTION: Has been said to contain 2 RING fingers, but these are  
 CC not found by any domain detection methods.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).

EMBL: AF343574; AAL73348.1; -  
 DR EMBL: AF381277; AAM21452.1; -  
 DR EMBL: AF381278; AAM21452.1; -  
 DR EMBL: AF381279; AAM21454.1; -  
 DR EMBL: AF381280; AAM21455.1; -  
 DR EMBL: AF381281; AAM21456.1; -  
 DR EMBL: AF168004; AAF34874.1; -  
 DR EMBL: AF210434; AAG37013.1; -  
 DR EMBL: AF257234; AAF68666.1; -  
 DR EMBL: AB039878; BAA92431.1; -  
 DR HSP: Q9W56; IMG8.  
 DR RGD: 61797; Park2.  
 DR InterPro: IPR003977; Parkin.  
 DR InterPro: IPR00626; Ubiquitin.  
 DR InterPro: IPR002867; Znf\_C6HC.  
 DR Pfam: PF01485; IBR: 1.  
 DR Pfam: PF00240; ubiquitin: 1.  
 DR PRINTS: PR01475; PARKIN.  
 DR PRINTS: PR00348; Ubiquitin.  
 DR SMART: SM00647; IBR: 2.  
 DR SMART: SM00213; UBO: 1.  
 DR PROSITE: PS00299; Ubiquitin\_1; FALSE\_NRG.  
 DR PROSITE: PS00053; Ubiquitin\_2; 1.  
 DR Alternative splicing; Ligase; Metal-binding; Repeat; S-nitrosylation;  
 KM Ub1 conjugation; Ub1 conjugation pathway; Zinc; Zinc-finger.  
 FT DOMAIN 1 204 238 Ubiquitin-like.  
 FT DOMAIN 257 293 STY1 binding 1.  
 FT ZN\_FING 313 377 STY1 binding 2.  
 FT ZN\_FING 401 457 IBR-type 1.  
 FT VARSPLIC 1 191 IBR-type 2.  
 FT VARSPLIC 57 57 Missing (in isoform 3 and isoform 6).  
 FT VARSPLIC 179 206 /FTId=VSP 011717.  
 FT VARSPLIC 390 394 /FTId=VSP 011719.  
 FT VARSPLIC 395 465 /FTId=VSP 011720.  
 FT VARSPLIC 429 446 /FTId=VSP 011721.  
 FT VARSPLIC 447 465 /FTId=VSP 011722.  
 FT CONFLICT 24 24 F -> C (in Ref. 1; AAF34874).  
 FT CONFLICT 138 138 E -> A (in Ref. 3).  
 FT CONFLICT 348 348 K -> R (in Ref. 3).  
 SQ SEQUENCE 465 AA; 51708 MW; E13CF170AD6D042B CRC64;  
 Query Match 85.4%; Score 2218; DB 1; Length 465;  
 Best Local Similarity 84.9%; Pred. No. 6,8e-166;  
 Matches 395; Conservative 26; Mismatches 44; Indels 0; Gaps 0;  
 QY 1 MIVFRENSSHGPEVVDSTISIFOLKEVAKRGVDPADQLRVIFAGKELENDWTYONCD 60  
 DB 1 MIVFRENSSHGPEVVDSTISIFOLKEVAKRGVDPADQLRVIFAGKELENDWTYONCD 60  
 QY 61 LDQGSIVHIVORPMKQGEENATGDDPRNAAAGCEBEPQSLRTVLDSSVLPEDSVGLA 120  
 DB 61 LDQGSIVHIVORPMKQGEENATGDDPRNAAAGCEBEPQSLRTVLDSSVLPEDSVGLA 120  
 QY 121 VILHTDSKDSPPAGSPAGRSIVNSFYVYCKGRCQAVQPGKLRYQGSTQCAATLTLTQGP 180  
 DB 121 VILHTDSKDSPPAGSPAGRSIVNSFYVYCKGRCQAVQPGKLRYQGSTQCAATLTLTQGP 180  
 QY 121 VILHTDSKDSPPAGSPAGRSIVNSFYVYCKGRCQAVQPGKLRYQGSTQCAATLTLTQGP 180  
 DB 121 VILHTDSKDSPPAGSPAGRSIVNSFYVYCKGRCQAVQPGKLRYQGSTQCAATLTLTQGP 180

QY 181 SCWDVLIIPNRSGECQSPHCPGTSAEFFKCGAHPSTSKETPVALH.IATNSRMITCIT 240  
 DB 181 SCWDVLIIPNRSGECQSPHCPGTSAEFFKCGAHPSTSKETPVALH.IATNSRMITCIT 240  
 QY 241 CTDRSPVLPVQCNHSHVYCLDCPFHYCYTRLRNDQFVHPDQGLSPCVAGCPNSL.IKE 300  
 DB 241 CTDRSPVLPVQCNHSHVYCLDCPFHYCYTRLRNDQFVHPDQGLSPCVAGCPNSL.IKE 300  
 QY 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPGCAGLILPEPDQKMTCEGNGIGC 360  
 DB 301 LHHFRILGEEQYNNRYQYGAEECVLQMGVLCPRPGCAGLILPEPDQKMTCEGNGIGC 360  
 QY 361 GFAFRCCKEATYHEEGCAVFEASGTTQAVRVDERAAQARWEAASKETIKKTKPCPR 420  
 DB 361 GFAFRCCKEATYHEEGCAVFEASGTTQAVRVDERAAQARWEAASKETIKKTKPCPR 420  
 QY 421 CHAVPEKNGGCMHMKCPQPOCRLECMNGGCEMNRVCMDHWEPU 465  
 DB 421 CHAVPEKNGGCMHMKCPQPOCRLECMNGGCEMNRVCMDHWEPU 465  
 RESULT 3  
 ID PRKN MOUSE STANDARD; PRT: 464 AA.  
 AC Q9W56; Q9ES22; Q9ES23;  
 DT 25-OCT-2004 (Rel. 45, Created)  
 DT 25-OCT-2004 (Rel. 45, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Parkin (EC 6.3.2.-) (Ubiquitin E3 ligase PRKN).  
 GN Name=Park2; Synonym=Prkn;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND SUBCELLULAR  
 RP LOCATION.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=20279843; PubMed=10818204;  
 RA Kitada T., Asakawa S., Minoshima S., Mizuno Y., Shimizu N.;  
 RT "Molecular cloning, gene expression, and identification of a splicing  
 RL variant of the mouse parkin gene.";  
 RL Mamm. Genome 11:417-421(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), TISSUE SPECIFICITY, AND  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=BALB/C; TISSUE=Kidney;  
 RX MEDLINE=20571394; PubMed=11122330;  
 RA Stiechel C.C., Augustin M., Kuehn K., Zhu X.-R., Engels P., Ullmer C.,  
 RA Luebbert H.;  
 RT "Parkin expression in the adult mouse brain.";  
 RL Eur. J. Neurosci. 12:4181-4194(2000).  
 RN [3]  
 RP TISSUE SPECIFICITY AND DEVELOPMENT.  
 RX PubMed=11675120;  
 RA Huynh D.P., Dy M., Nguyen D., Kiehl T.-R., Pulst S.M.;  
 RT "Differential expression and tissue distribution of parkin isoforms  
 RL during mouse development.";  
 RL Brain Res. Dev. Brain Res. 130:173-181(2001).  
 RN [4]  
 RP FUNCTION IN UbiquITINATION AND S-NITROSylation.  
 RX PubMed=15105460; DOI=10.1126/science.1093891;  
 RA Chung K.K., Thomas B., Li X., Pletnikova O., Troncoso J.C., Marsh L.,  
 RA Dawson V.L., Dawson T.M.;  
 RT "S-nitrosylation of parkin regulates ubiquitination and compromises  
 RL parkin's protective function.";  
 RL Science 304:1328-1331(2004).  
 RN [5]  
 RP STRUCTURE BY NMR OF 1-76.  
 RX PubMed=12652124;  
 RA Tashiro M., Okubo S., Shimotakahara S., Hatanaka H., Yasuda H.,  
 RA Kashiwano M., Yokoyama S., Shindo H.;  
 RT "NMR structure of ubiquitin-like domain in PARKIN: gene product of

familial Parkinson's disease." ;  
 CC J. Biolomol. NMR 25:153-156(2003).  
 CC -1- FUNCTION: Functions within a multiprotein E3 ubiquitin ligase  
 CC complex, catalyzing the covalent attachment of ubiquitin moieties  
 CC onto substrate proteins. These substrates include SYT11, CENBL,  
 CC GPR37, STUB1, a 22 kDa O-linked glycosylated isoform of SNCAIP and  
 CC SEPT5. May play a more general role in the ubiquitin proteasomal  
 CC pathway by participating in the removal and/or detoxification of  
 CC abnormally folded or damaged protein. Loss of this ubiquitin  
 CC ligase activity appears to be the mechanism underlying  
 CC pathogenesis of AR-OP. May protect neurons against alpha synuclein  
 CC toxicity, proteasomal dysfunction, GPR37 accumulation, and  
 CC kainate-induced excitotoxicity. May play a role in controlling  
 CC neurotransmitter trafficking at the presynaptic terminal and in  
 CC calcium-dependent exocytosis. Regulates cyclin E during neuronal  
 CC apoptosis. May represent a tumor suppressor gene (By similarity).  
 CC -1- PATHWAY: Ubiquitin conjugation; third step.  
 CC -1- SUBUNIT: Forms an E3 ubiquitin ligase complex with UBE2L3 or  
 CC UBE2L6. Part of a SCF-like complex, consisting of PARK2, CUL1 and  
 CC FBXW7. Interacts with SNCAIP. Binds to the C2A and C2B domains of  
 CC SYT11. Interacts and regulates the turnover of SEPT5. Part of a  
 CC complex, including STUB1, HSP70, and GPR37. The amount of STUB1 in  
 CC the complex increases during ER stress. STUB1 promotes the  
 CC dissociation of HSP70 from PARK2 and GPR37, thus facilitating  
 CC PARK2-mediated GPR37 ubiquitination. HSP70 transiently associates  
 CC with unfolded GPR37 and inhibits the E3 activity of PARK2,  
 CC whereas, STUB1 enhances the E3 activity of PARK2 through promotion  
 CC of dissociation of HSP70 from PARK2-GPR37 complexes. Interacts  
 CC with PSMD4 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Expressed in the endoplasmic  
 CC reticulum, dendrites, some presynaptic terminals and in  
 CC postsynaptic dendrites.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Name=1;  
 CC IsoId=Q9WVS6-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9WVS6-2; Sequence=VSP\_011714, VSP\_011715;  
 CC Name=3;  
 CC IsoId=Q9WVS6-3; Sequence=VSP\_011713, VSP\_011716;  
 CC -1- TISSUE SPECIFICITY: Expressed in all subdivisions of the brain.  
 CC Highly expressed in brainstem, cranial nerve, pontine, cerebellar  
 CC nuclei, indusium griseum, nuclei reticularis, strata optica and  
 CC lacunosum moleculare of the hippocampal CA2 region. Low levels  
 CC were found in the telencephalon and diencephalon. Expressed in  
 CC heart, liver, skeletal muscle, kidney and testis.  
 CC -1- DEVELOPMENTAL STAGE: In late 10 dpc weakly expressed in  
 CC postmitotic neurons in the mantle layer of the developing nervous  
 CC system. Expression increased at 11-12 dpc. At 15-16 dpc, as more  
 CC specialized neurons and nonneural cells are formed, expression is  
 CC more tissue specific. Expression was highest in the neurites,  
 CC moderate levels were observed in the migrating postmitotic neurons  
 CC in the intermediate and neopallial layers. In the diencephalon and  
 CC other CNS regions, while the weakest level of expression was  
 CC observed in the cell bodies. In nonneural tissues, high levels of  
 CC expression were found in the muscle walls of the intestine, the  
 CC blood vessels and the dermis.  
 CC -1- DOMAIN: The ubiquitin-like domain binds the PSMD4 subunit of 26S  
 CC proteasomes (By similarity).  
 CC -1- PTM: Auto-ubiquitinates in an E2-dependent manner leading to its  
 CC own degradation (By similarity).  
 CC -1- PTM: S-nitrosylated.  
 CC -1- SIMILARITY: Contains 2 IRR-type zinc fingers.  
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC -1- CAUTION: Has been said to contain 2 zinc fingers, but these are  
 CC not found by any domain detection methods.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>)

or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch).  
 CC -----  
 CC EMBL; AB019558; BAA82404.1; -  
 CC DR EMBL; AF250293; AAG13890.1; -  
 CC DR EMBL; AF250294; AAG13891.1; -  
 CC DR EMBL; AF250295; AAG13892.1; -  
 CC DR PDB; 1MG8; NMR; A=1-76.  
 CC MGD; MGI:1355296; Park2.  
 CC DR InterPro; IPR003977; parkin.  
 CC DR InterPro; IPR000626; Ubiquitin.  
 CC DR InterPro; IPR002867; ZnF\_C6HC.  
 CC DR Pfam; PF01485; IRR; 1.  
 CC DR Pfam; PF00240; ubiquitin; 1.  
 CC DR PRINTS; PR01475; PARKIN.  
 CC DR PRINTS; PR00348; UBIQUITIN.  
 CC DR SMART; SM00647; IRR; 2.  
 CC DR SMART; SM00213; UBO; 1.  
 CC DR PROSITE; PS00299; UBIQUITIN\_1; FALSE\_NEG.  
 CC DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 CC DR 3D-structure; Alternative splicing; Ligase; Metal-binding; Repeat;  
 CC KW S-nitrosylation; Ub1 conjugation; Ub1 conjugation pathway; Zinc;  
 CC Zinc-finger.  
 CC FT DOMAIN 1 76 Ubiquitin-like.  
 CC FT DOMAIN 204 238 SYT11 binding 1.  
 CC FT DOMAIN 257 293 SYT11 binding 2.  
 CC FT ZN\_FING 313 377 IRR-type 1.  
 CC FT ZN\_FING 400 456 IRR-type 2.  
 CC FT VARSPLIC 244 261 RSPVAVFOCNHNVICLD -> SHLPLSGASVWTRPHLH  
 CC (in isoform 3).  
 CC FT FTID=VSP\_011713.  
 CC FT VARSPLIC 245 254 SPVLVFOCNH -> FMRMSKRTS (in isoform 2).  
 CC FT FTID=VSP\_011714.  
 CC FT VARSPLIC 255 464 Missing (in isoform 2).  
 CC FT VARSPLIC 262 464 Missing (in isoform 3).  
 CC FT FTID=VSP\_011716.  
 CC FT CONFLICT 137 137 P -> PA (in Ref. 1 and 2; AAG13890).  
 CC SQ SEQUENCE 464 AA; 51617 MW; 5574A285A9A1B080 CRC64;  
 CC -----  
 CC Query Match 83.6%; Score 2170.5; DB 1; Length 464;  
 CC Best Local Similarity 83.4%; Pred. No. 3-6e-162;  
 CC Matches 388; Conservative 29; Mismatches 47; Indels 1; Gaps 1;  
 CC 1 MIVFVRPNSSHGFPPEVNSDTSIFOLKEVYVARGVAPDQLRVIFAKGLRNDMTVQNC 60  
 CC 1 MIVFVRPNSSYGFPEVNSDTSILQKEVYVARGVAPDQLRVIFAKGLRNDMTVQNC 60  
 CC 61 LDQOSIVHIVQRPWRKQGMNATGDDPRNNAAGCEREPOSILRVDSLSSVLPQDSVGLA 120  
 CC 61 LEQOSIVHIVQRPWRKSHETNATSGDDPQSTSEGSIMESRSILRVDSLSSHTLPVDSVGLA 120  
 CC 121 VILHFDKSRKDEPPAGSPAGRSITNSFFYYCKPCQRPVQPGKRLRVQGSTORNTLTITQGP 180  
 CC 121 VILHFDKSRKDEPPAGSPAGRSITNSFFYYCKPCQRPVQPGKRLRVQGSTORNTLTITQGP 180  
 CC 121 VILHFDKSRKDEPPAGSPAGRSITNSFFYYCKPCQRPVQPGKRLRVQGSTORNTLTITQGP 180  
 CC 121 VILHFDKSRKDEPPAGSPAGRSITNSFFYYCKPCQRPVQPGKRLRVQGSTORNTLTITQGP 180  
 CC 181 SCMDVLLIPNMRSGCQSPHCGTSAEFKCGAHPSTKXENPVLLHLATNSRNTCT 240  
 CC 181 SCMDVLLIPNMRSGCQSPHCGTSAEFKCGAHPSTKXENPVLLHLATNSRNTCT 240  
 CC 180 SCMDVLLIPNMRSGCQSPHCGTSAEFKCGAHPSTKXENPVLLHLATNSRNTCT 239  
 CC 241 CTDVASPVLPFOCNRRHYILCDFHLYCVTRLRNDQFVNDPLQGYSLPCVACGPNLSLKE 300  
 CC 240 CTDVASPVLPFOCNRRHYILCDFHLYCVTRLRNDQFVNDPLQGYSLPCVACGPNLSLKE 299  
 CC 301 LHHFRILGEEQYRNYQYGAEECVTOMGCVLCPRRCCGAGLLPEPDQRYRTVEGGNGLCG 360  
 CC 300 LHHFRILGEEQYRNYQYGAEECVTOMGCVLCPRRCCGAGLLPEPDQRYRTVEGGNGLCG 359  
 CC 361 GPAFRCKEAEHRECEGSAVFEAGSTTQAVYVDRBAEQAWEAASKETIKTKTPCPR 420  
 CC 360 GPAFRCKEAEHRECEGSAVFEAGSTTQAVYVDRBAEQAWEAASKETIKTKTPCPR 419  
 CC 421 CHVPEKNGKGMKMKPOPOCRLEWCNMGCEMNVRNVMGDHMFV 465  
 CC 420 CHVPEKNGKGMKMKPOPOCRLEWCNMGCEMNVRNVMGDHMFV 459

Db 420 CNVPIKXNGCMHMKCPQCKLEMCWNCGEMNRACWGDHWPDU 464

RESULT 4

Q8N142 PRELIMINARY; PRT; 203 AA.

AC Q8N142; 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)

DE Parkin isoform.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

OK NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA D'Agata V., Scapagnini G., Cavallaro S.; Submitted (May-2001) to the EMBL/GenBank/DBJ databases.

RL EMBL: AF381284; AAM21459.1; -

DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.

DR GO; GO:0006512; P:ubiquitin cycle; IEA.

DR InterPro; IPR003977; parkin.

DR InterPro; IPR002867; Znf\_C6HC.

DR Pfam; PF01485; IIR; 1

DR PRINTS; PRO1475; PARKIN.

DR SMART; SM00647; IIR; 1.

DR SMART; SM00647; IIR; 1.

SQ SEQUENCE 203 AA; 22192 MW; F92E796F069DC818 CRC64;

Query Match 43.0%; Score 1115; DB 2; Length 203;

Best Local Similarity 100.0%; Pred. No. 1,2e-79;

Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 MSGECSPHCPTGSAFFKCGAHPTSDKETPVALLIATNSRNITCITCDVRSPLVLF 251

Db 1 MSGECSPHCPTGSAFFKCGAHPTSDKETPVALLIATNSRNITCITCDVRSPLVLF 60

QY 252 QCNRRVILGDCPHLYCYTRLNDROFVHPDPOGYSIPCYAGCPNLIKELHFRILGEEQ 311

Db 61 QCNRRVILGDCPHLYCYTRLNDROFVHPDPOGYSIPCYAGCPNLIKELHFRILGEEQ 120

QY 312 YNRQOQYGAEECVLMQGVLCPRPGCAGLIPERPOKRYTCGGNGLGGFAPCECKEA 371

Db 121 YNRQOQYGAEECVLMQGVLCPRPGCAGLIPERPOKRYTCGGNGLGGFAPCECKEA 180

QY 372 YHSGCSAVFEASGTTTQ 389

Db 181 YHSGCSAVFEASGTTTQ 198

RESULT 5

Q95M07 PRELIMINARY; PRT; 194 AA.

AC Q95M07; 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)

DE Parkin (Fragment).

GN Name-parkin gene.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

OC NCBI\_TaxID=9913;

OK NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RA TISSUE=Brain;

RC MEDLINE=98219084; PubMed=9560156; DOI=10.1038/33416;

RA Kitada T., Asakawa S., Hattori N., Matsumine H., Yamamura Y., Minoshima S., Yokochi M., Mizuno Y., Shimizu N.; "Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.";

RT Nature 392:605-608 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=21869006; PubMed=11879804; DOI=10.1016/S0006-8993(02)02241-2;

RA Hase A., Yamada H., Arai K., Sunada Y., Shimizu T., Matsumura K.; "Characterization of parkin in bovine peripheral nerve.";

RL Brain Res. 930:143-149 (2002).

DR EMBL; AB060701; BAB70670.1; -

DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.

DR GO; GO:0006512; P:ubiquitin cycle; IEA.

DR InterPro; IPR003977; parkin.

DR PRINTS; PRO1475; PARKIN.

DR NON\_TER 1 1

FT NON\_TER 194 194

SQ SEQUENCE 194 AA; 21225 MW; 9B214F09FA74FAB1 CRC64;

Query Match 40.8%; Score 1060; DB 2; Length 194;

Best Local Similarity 99.5%; Pred. No. 2,4e-75;

Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 92 AGGCRERQSLTRVDLSSVLPQDSVGLAVILHTDSKSDPSPAGSGRSTYNSFYVYCK 151

Db 1 AGGCRERQSLTRVDLSSVLPQDSVGLAVILHTDSKSDPSPAGSGRSTYNSFYVYCK 60

QY 152 GPCQRPVQPKLRVOCSTCRQATLTLTQOPSCMDVLIIPNMSGCGSPHCGTSAEFFK 211

Db 61 GPCQRPVQPKLRVOCSTCRQATLTLTQOPSCMDVLIIPNMSGCGSPHCGTSAEFFK 120

QY 212 CGAHPTSDKETPVALLIATNSRNITCITCDVRSPLVLFQCNRRVILGDCPHLYCYTRL 271

Db 121 CGAHPTSDKETPVALLIATNSRNITCITCDVRSPLVLFQCNRRVILGDCPHLYCYTRL 180

QY 272 LNDROFVHPDPOGYSIPCYAGCPNLIKELHFRILGEEQ 285

Db 181 LNDROFVHPDPOGYSIPCYAGCPNLIKELHFRILGEEQ 194

RESULT 6

Q95T14 PRELIMINARY; PRT; 482 AA.

AC Q95T14; 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE SD01679P.

GN Name-park; Synonyms=CG10523;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.

OC NCBI\_TaxID=7227;

OK NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Strapleton M., Brokstein P., Hong L., Agdayani A., Carlson J., Chapple M., Chavez C., Dorsett V., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M., Celniker S.; Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY058754; AAL13983.1; -

DR HSSP; Q862M4; 1AAR.

DR FlyBase; FBgn0041100; park.

DR GO; GO:0007005; P:mitochondrion organization and biogenesis; IMP.

DR InterPro; IPR002052; N6 Mtase.

DR InterPro; IPR003977; parkin.

DR InterPro; IPR000626; Ubiquitin.

DR InterPro; IPR002867; Znf\_C6HC.

DR Pfam; PF01485; IIR; 1.

DR Pfam; PF00240; ubiquitin; 1.

DR PRINTS; PRO1475; PARKIN.

DR PRINTS; PRO0348; UBIQUITIN.

DR SMART; SM00647; IIR; 2.

DR SMART; SM00213; UBQ; 1.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 25, 2005, 16:48:21 ; Search time 2103.69 Seconds  
(without alignments)  
10848.741 Million cell updates/sec

Title: US-10-622-817-10  
471  
Perfect score: 1 ggaagccagcagcagtagatc.....accctcgtcctgtcgtag 471  
Sequence:

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_bcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sbs:\*  
12: gb\_gy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	471	100.0	471	6	AX224425 Sequence
2	468	99.4	2960	6	BD269536 Mutations
3	466.4	99.0	904	6	AF381283 Homo sapi
4	466.4	99.0	2960	6	AR492172 Sequence
5	466.4	99.0	2960	6	AX584224 Sequence
6	466.4	99.0	2960	6	AX584226 Sequence
7	466.4	99.0	2960	6	AB009973 Homo sapi
8	464.8	98.7	1575	9	BC022014 Homo sapi
9	460.8	97.8	1157	9	AF381282 Homo sapi
10	454	96.4	583	4	AB060701 Bos tauru
11	353.4	75.0	1564	10	AB039878 Rattus no
12	350.2	74.4	904	10	AF343575 Rattus no
13	348.6	74.0	1466	10	AF168004 Rattus no
14	348.6	74.0	1576	10	AF210434 Rattus no
15	348.2	73.9	1157	10	AF343574 Rattus no
16	347	73.7	1398	10	AF257234 Rattus no
17	347	73.7	1470	10	AF381277 Rattus no
18	347	73.7	1554	10	AF381280 Rattus no
19	347	73.7	1578	10	AF381278 Rattus no

20	340.2	72.2	1644	10	AB019558 Mus muscu
21	340.2	72.2	3043	6	AX089875 Sequence
22	340.2	72.2	3043	6	AX136840 Sequence
23	340.2	72.2	3092	6	AX089879 Sequence
24	340.2	72.2	3092	6	AX136844 Sequence
25	340.2	72.2	3226	10	AF250293 Mus muscu
26	340.2	72.2	3253	6	AX089876 Sequence
27	340.2	72.2	3253	6	AX089878 Sequence
28	340.2	72.2	3253	6	AX136841 Sequence
29	340.2	72.2	3253	6	AX136843 Sequence
30	340.2	72.2	3254	6	AX089877 Sequence
31	340.2	72.2	3254	6	AX136842 Sequence
32	340.2	72.2	3255	6	AX089863 Sequence
33	340.2	72.2	3255	6	AX089880 Sequence
34	340.2	72.2	3255	6	AX089881 Sequence
35	340.2	72.2	3255	6	AX089882 Sequence
36	340.2	72.2	3255	6	AX136828 Sequence
37	340.2	72.2	3255	6	AX136845 Sequence
38	340.2	72.2	3255	6	AX136846 Sequence
39	340.2	72.2	3255	6	AX136847 Sequence
40	338.6	71.9	3014	6	AX089869 Sequence
41	338.6	71.9	3014	6	AX136834 Sequence
42	338.4	71.8	1158	9	AF381284 Homo sapi
43	288.4	61.2	2876	6	AR492173 Sequence
44	265.4	56.3	3136	6	AX089872 Sequence
45	265.4	56.3	3136	6	AX136837 Sequence

ALIGNMENTS

RESULT 1  
LOCUS AX224425 471 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 3 from Patent WO0160857.  
ACCESSION AX224425  
VERSION AX224425.1 GI:15554665  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Koultnikova,H., Brice,A., Fournier,A., Pradier,L., Prades,C.,  
Arnould-Reguigne,I., Rosier-Montus,M.F. and Corti,O.  
TITLE Compositions useful for regulating parkin gene activity  
JOURNAL Patent: WO 0160857-A 3 23-AUG-2001;  
Aventis Pharma S.A. (FR) ; INSTITUT NATIONAL DE LA SANTE ET DE LA  
RECHERCHE MEDICALE (INSERM) (FR)

FEATURES  
source  
location/Qualifiers  
1..471  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
<1..471  
/note="unamed protein product"

CDS  
/codon\_start=1  
/protein\_id="CAC69572.1"  
/db\_xref="GI:15554666"  
/translation="GSPDGRSIVNSFYVCKGPCQVOPGRLRVQSTCRQATLLTQ  
GPSQMDVLIPNRMSGECOSPHCGTSAEFCGAPTSKETSVALHLATNSRNI  
TCITCTIVRSPLVIFQNSRHVICLDGFHLVCTVRINDRQVHDPQGLSLPCV"

ORIGIN  
Query Match 100.0%; Score 471; DB 6; Length 471;  
Best Local Similarity 100.0%; Pred. No. 2,3e-143;  
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGAAGTCAGCAGGATGATCAATCTTACCAACGCTTTATGTGTTGCAAGGCCCTGT 60  
|||||  
DB 1 GGAAGTCAGCAGGATGATCAATCTTACCAACGCTTTATGTGTTGCAAGGCCCTGT 60  
|||||  
OY 61 CAAGAGTCAGCGGGAACACTCAGGATACAGTGCAGACCTGCAGGCAAGCAAGCTC 120

Db 61 CAAAGAGTGACGCGGAAAACTCAGGGTACAGTGACAGACCTGACAGGACGCTC 120  
Qy 121 ACCGTGACCCAGGGTCCATCTTGCTGGAGATGATTTTAATCCAAACCGATGAGTGT 180  
Db 121 ACCTTGACCCAGGGTCCATCTTGCTGGAGATGATTTTAATCCAAACCGATGAGTGT 180  
Qy 181 GAATGCCAATCCCACTGCGCTGGAGTGTGAGAGATTTTCTTTAAATGTGAGCA 240  
Db 181 GAATGCCAATCCCACTGCGCTGGAGTGTGAGAGATTTTCTTTAAATGTGAGCA 240  
Qy 241 CACCCCACTCTGACAAAGAAACATCAGTCTTTGACCTGATCGCAAAATAGTCG 300  
Db 241 CACCCCACTCTGACAAAGAAACATCAGTCTTTGACCTGATCGCAAAATAGTCG 300  
Qy 301 AACATCATTGACATTAAGTACAGACAGTGTGAGAGCCGCTCGTGTTCAGTGCAAC 360  
Db 301 AACATCATTGACATTAAGTACAGACAGTGTGAGAGCCGCTCGTGTTCAGTGCAAC 360  
Qy 361 TCCCGCAGGTGATTTGCTTAGACTGTTCACCTTAAGTGTGACAAAGACTCAATGAT 420  
Db 361 TCCCGCAGGTGATTTGCTTAGACTGTTCACCTTAAGTGTGACAAAGACTCAATGAT 420  
Qy 421 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 471  
Db 421 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 471

RESULT 2  
BD269536 2960 bp DNA linear PAT 17-JUL-2003  
LOCUS Mutations of the parkin gene, compositions, methods and uses.  
DEFINITION BD269536  
VERSION BD269536.1 GI:33079304  
KEYWORDS JP 2002536961-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2960)  
Bouley, S.  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Bouley, S., Lucking, C., Abbas, N.E., Deneffe, P., Ricard, S. and

TITLE Mutations of the parkin gene, compositions, methods and uses  
JOURNAL AVENTIS PHARMA SA, INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE  
MEDICALE

COMMENT OS Homo sapiens (human)  
PN JP 2002536961-A/1  
PD 05-NOV-2002  
PF 18-NOV-1999 JP 2000544062  
PR 19-NOV-1998 FR 98/14524, 12-MAR-1999 US 60/124239 PR  
PI ALEXIS BRICE, CHRISTOPHE LUCKING, NANCER EDDINE ABBAS, PATRICE PI  
DENEFFE,  
PI SYLVAIN RICARD, SANDRINE BOULEY  
PC C12N1/21,  
PC C12N5/10, C12Q1/68, G01N33/15, G01N33/50, C12N5/00, C12N5/00 CC  
Mutations of the parkin gene, compositions, methods and uses FH  
Key Location/Qualifiers  
FT source 1..2960  
FT location/Qualifiers  
FT /organism='Homo sapiens (human)'.  
FT 1..2960

FEATURES  
source  
1..2960  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 99.4%; Score 468; DB 6; Length 2960;  
Best Local Similarity 100.0%; Pred. No. 2.9e-142;  
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAAGTCCAGAGGTAGATCAATCTACAAACAGCTTTTATGTGTATTCGAAAGGCCCTGT 60  
Db 504 GGAAGTCCAGAGGTAGATCAATCTACAAACAGCTTTTATGTGTATTCGAAAGGCCCTGT 563  
Qy 61 CAAAGAGTGACGCGGAAAACTCAGGGTACAGTGACAGACCTGACAGGACGCTC 120  
Db 564 CAAAGAGTGACGCGGAAAACTCAGGGTACAGTGACAGACCTGACAGGACGCTC 120  
Qy 121 ACCTTGACCCAGGGTCCATCTTGCTGGAGATGATTTTAATCCAAACCGATGAGTGT 180  
Db 121 ACCTTGACCCAGGGTCCATCTTGCTGGAGATGATTTTAATCCAAACCGATGAGTGT 180  
Qy 181 GAATGCCAATCCCACTGCGCTGGAGTGTGAGAGATTTTCTTTAAATGTGAGCA 240  
Db 181 GAATGCCAATCCCACTGCGCTGGAGTGTGAGAGATTTTCTTTAAATGTGAGCA 240  
Qy 241 CACCCCACTCTGACAAAGAAACATCAGTCTTTGACCTGATCGCAAAATAGTCG 300  
Db 241 CACCCCACTCTGACAAAGAAACATCAGTCTTTGACCTGATCGCAAAATAGTCG 300  
Qy 301 AACATCATTGACATTAAGTACAGACAGTGTGAGAGCCGCTCGTGTTCAGTGCAAC 360  
Db 301 AACATCATTGACATTAAGTACAGACAGTGTGAGAGCCGCTCGTGTTCAGTGCAAC 360  
Qy 361 TCCCGCAGGTGATTTGCTTAGACTGTTCACCTTAAGTGTGACAAAGACTCAATGAT 420  
Db 361 TCCCGCAGGTGATTTGCTTAGACTGTTCACCTTAAGTGTGACAAAGACTCAATGAT 420  
Qy 421 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 468  
Db 421 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 468  
Qy 924 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 971  
Db 924 CGGCACTTTGTTACAGACCTGCACTTGCTACTCCCTGCTTGTGTGAG 971

RESULT 3  
AF381283 904 bp mRNA linear PRI 02-MAY-2002  
LOCUS Homo sapiens parkin isoform mRNA, complete cds, alternatively  
DEFINITION AF381283  
ACCESSION AF381283  
VERSION AF381283.1 GI:20385799  
KEYWORDS

REFERENCE 1 (bases 1 to 904)  
Bouley, S.  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Bouley, S., Lucking, C., Abbas, N.E., Deneffe, P., Ricard, S. and

TITLE Mutations of the parkin gene, compositions, methods and uses  
JOURNAL AVENTIS PHARMA SA, INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE  
MEDICALE

COMMENT OS Homo sapiens (human)  
PN JP 2002536961-A/1  
PD 05-NOV-2002  
PF 18-NOV-1999 JP 2000544062  
PR 19-NOV-1998 FR 98/14524, 12-MAR-1999 US 60/124239 PR  
PI ALEXIS BRICE, CHRISTOPHE LUCKING, NANCER EDDINE ABBAS, PATRICE PI  
DENEFFE,  
PI SYLVAIN RICARD, SANDRINE BOULEY  
PC C12N1/21,  
PC C12N5/10, C12Q1/68, G01N33/15, G01N33/50, C12N5/00, C12N5/00 CC  
Mutations of the parkin gene, compositions, methods and uses FH  
Key Location/Qualifiers  
FT source 1..904  
FT location/Qualifiers  
FT /organism="Homo sapiens"  
FT /mol\_type="mRNA"  
FT /db\_xref="taxon:9606"  
FT 1..904  
FT /note="alternatively spliced; transcript variant 7"  
FT /codon\_start=1  
FT /product="parkin isoform"  
FT /protein\_id="P44714.8.1"  
FT /db\_xref="GI:20385800"  
FT /translation="MNATGDDPRNAAAGCERBPQSLTRVDLSSVLPQDSVGLAVIL  
HTDSRKDPSPPAGSPAGRSIYNSFYVCKGQVQKRLKVCSTROATLITLQGPS  
CWDVILIPNMSGECOSPHCPGTSAEPFGCAHPSDKETPVAILHILITNSNITCI  
TCDVRSPLVIFQCNRSRHYCLDPCFLHYCTRLNDRQFVHDPQLGSLPCVVCLLRGM"

FEATURES  
source  
1..904  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
74..730  
/note="alternatively spliced; transcript variant 7"  
/codon\_start=1  
/product="parkin isoform"  
/protein\_id="P44714.8.1"  
/db\_xref="GI:20385800"  
/translation="MNATGDDPRNAAAGCERBPQSLTRVDLSSVLPQDSVGLAVIL  
HTDSRKDPSPPAGSPAGRSIYNSFYVCKGQVQKRLKVCSTROATLITLQGPS  
CWDVILIPNMSGECOSPHCPGTSAEPFGCAHPSDKETPVAILHILITNSNITCI  
TCDVRSPLVIFQCNRSRHYCLDPCFLHYCTRLNDRQFVHDPQLGSLPCVVCLLRGM"

ORIGIN



Query Match 99.0%; Score 466.4; DB 9; Length 904;  
Best Local Similarity 99.8%; Pred. No. 8.2e-142;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 60  
DB 239 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 298  
QY 61 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 120  
DB 299 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 358  
QY 121 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 180  
DB 359 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 418  
QY 181 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
DB 419 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 478  
QY 241 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCGCAACAATAGTCGG 300  
DB 479 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCGCAACAATAGTCGG 538  
QY 301 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 360  
DB 539 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 598  
QY 361 TCCCGCAGCTGATTTGCTTAGACTGTTTCCATTATACGTGTGACAAAGATCAATGAT 420  
DB 599 TCCCGCAGCTGATTTGCTTAGACTGTTTCCATTATACGTGTGACAAAGATCAATGAT 658  
QY 421 CGGAGTTTGTTCACGACACCTCACTTGGCTACCTCCCTGCTGTGTG 468  
DB 659 CGGAGTTTGTTCACGACACCTCACTTGGCTACCTCCCTGCTGTGTG 706

RESULT 4  
AR492172  
LOCUS AR492172 2960 bp DNA linear PAT 15-MAY-2004  
DEFINITION Sequence 1 from patent US 6716621.  
ACCESSION AR492172  
VERSION AR492172.1 GI:47260688  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2960)  
AUTHORS Shimizu N. and Mizuno Y.  
TITLE Isolated DNA or gene responsible for Parkinson's disease  
JOURNAL Patent: US 6716621-A 1 06-APR-2004;  
FEATURES  
LOCATION/Qualifiers  
SOURCE 1..2960  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 99.0%; Score 466.4; DB 6; Length 2960;  
Best Local Similarity 99.8%; Pred. No. 9.6e-142;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 60  
DB 504 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 563  
QY 61 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 120  
DB 564 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 623  
QY 121 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 180  
DB 624 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 683

QY 181 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
DB 684 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 743  
QY 241 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCCGAACAAATAGTCGG 300  
DB 744 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCCGAACAAATAGTCGG 803  
QY 301 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 360  
DB 804 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 863  
QY 361 TCCCGCAGCTGATTTGCTTAGACTGTTTCCATTATACGTGTGACAAAGATCAATGAT 420  
DB 864 TCCCGCAGCTGATTTGCTTAGACTGTTTCCATTATACGTGTGACAAAGATCAATGAT 923  
QY 421 CGGAGTTTGTTCACGACACCTCACTTGGCTACCTCCCTGCTGTGTG 468  
DB 924 CGGAGTTTGTTCACGACACCTCACTTGGCTACCTCCCTGCTGTGTG 971

RESULT 5  
AX584224  
LOCUS AX584224 2960 bp DNA linear PAT 10-JAN-2003  
DEFINITION Sequence 1 from Patent WO02079459.  
ACCESSION AX584224  
VERSION AX584224.1 GI:27655742  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Jensen, P. H.  
TITLE Means for inhibiting proteolytical processing of parkin  
JOURNAL Patent: WO 02079459-A 1 10-OCT-2002;  
FEATURES  
LOCATION/Qualifiers  
SOURCE 1..2960  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
mRNA  
Query Match 99.0%; Score 466.4; DB 6; Length 2960;  
Best Local Similarity 99.8%; Pred. No. 9.6e-142;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 60  
DB 504 GGAAGTCACAGAGTAGATCAATCTACACAGCTTTTATGTGTATTTGGAAAGGCCCTGT 563  
QY 61 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 120  
DB 564 CAAAGAGTGCAGCGGGGAAAACTCAGGGTACAGTGCAGACCTGCAGGCAAGGCAACGCTC 623  
QY 121 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 180  
DB 624 ACCTTGACCCAGAGGTCATCTTGCTGGAGTAGTATTTTAAATCCAAACCGGATGATGCT 683  
QY 181 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
DB 684 GAATGCCAATCCCACTGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 743  
QY 241 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCCGAACAAATAGTCGG 300  
DB 744 CACCCCACTCTGACAAAGAAACATCACTAGCTTTGACCTGATCCGAACAAATAGTCGG 803  
QY 301 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 360  
DB 804 AACATCACTTGACATTAAGTGCAGACAGTGCAGAGCCCGCTGCTGTTTCCAGTGCAC 863

Oy		361	TCCGCCACGCGATTGGTGTAGACTCTGTTTCACCTTAATACCTGTGTGACAGACTCAATGAT	420
Db		864	TCCGCCACGCGATTGGTGTAGACTCTGTTTCACCTTAATACCTGTGTGACAGACTCAATGAT	923
Oy		421	CGGCAGTTTGTTTCACGACCCTCACTTAGCTAGCTCCTGCCTGTGTG	468
Db		924	CGGCAGTTTGTTTCACGACCCTCACTTAGCTAGCTCCTGCCTGTGTG	971

  

RESULT 6		AX584226		
LOCUS	AX584226	2960 bp	DNA	linear
DEFINITION	Sequence 3 from Patent WO2079459.			PAT 10-JUN-2002
ACCESSION	AX584226			
VERSION	AX584226.1	GI:27655743		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
JOURNAL	Jensen, P.H. Means for inhibiting proteolytical processing of parkin Patent: WO 02079459-A 3 10-OCT-2002; NGENE A/S (DK)			
FEATURES	Location/Qualifiers			
source	1..2960 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"			
ORIGIN	mRNA 1..2960			

  

Query Match	99.0%;	Score 466.4;	DB 6;	Length 2960;
Best Local Similarity	99.8%;	Pred. No. 9.6e-142;		
Matches 467;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps				0;

  

Oy		1	GGAGGTCCAGCGAGTATCATCTTACAACAGCTTTAATGTGATTTGCAAGGCCCTGT	60
Db		504	GGAGGTCCAGCGAGTATCATCTTACAACAGCTTTAATGTGATTTGCAAGGCCCTGT	563
Oy		61	CMAAGGTGACGCCGGGAAACTCAGGGTACAGTGCAGACCTGCGAGGAGGCAAGCCTC	120
Db		564	CMAAGGTGACGCCGGGAAACTCAGGGTACAGTGCAGACCTGCGAGGAGGCAAGCCTC	623
Oy		121	ACCTTGACCCAGGGTCCATCTTGTCTGGGATATGTTTTAATTCGAAAACCGATAGTGT	180
Db		624	ACCTTGACCCAGGGTCCATCTTGTCTGGGATATGTTTTAATTCGAAAACCGATAGTGT	683
Oy		181	GAATGCCAATCCCCACATGCCCCTGGGGCTAGTGCAGAATTTTCTTAAATGTGGACA	240
Db		684	GAATGCCAATCCCCACATGCCCCTGGGGCTAGTGCAGAATTTTCTTAAATGTGGACA	743
Oy		241	CACCCACCTTGACAGAGAAACATCAATGATCTTTGCACTGATCGCACAAATAGTCG	300
Db		744	CACCCACCTTGACAGAGAAACATCAATGATCTTTGCACTGATCGCACAAATAGTCG	803
Oy		301	AACATCACTTSCATTAAGTGCAGAGTACAGAGCCCGCTGCTGTTTCCAGTCAAC	360
Db		804	AACATCACTTSCATTAAGTGCAGAGTACAGAGCCCGCTGCTGTTTCCAGTCAAC	863
Oy		361	TCCGCCACGCGATTGGTGTAGACTCTGTTTCACCTTAATACCTGTGTGACAGACTCAATGAT	420
Db		864	TCCGCCACGCGATTGGTGTAGACTCTGTTTCACCTTAATACCTGTGTGACAGACTCAATGAT	923
Oy		421	CGGCAGTTTGTTTCACGACCCTCACTTAGCTAGCTCCTGCCTGTGTG	468
Db		924	CGGCAGTTTGTTTCACGACCCTCACTTAGCTAGCTCCTGCCTGTGTG	971

  

RESULT 7	AB009973	2960 bp	mRNA	linear	FBI 14-APR-2000
LOCUS	AB009973				

DEFINITION	Homo sapiens mRNA for Parkin, complete cds.
ACCESSION	AB009973
VERSION	AB009973.1 GI:3063387
KEYWORDS	Parkin.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (sites)
TITLE	Kitada,T., Asakawa,S., Hattori,N., Matsumine,H., Yamamura,Y., Minoshima,S., Yokochi,M., Mizuno,Y. and Shimizu,N.
JOURNAL	Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism
MEDLINE	Nature 392 (6676), 605-608 (1998)
PIUMED	98219084
REFERENCE	9560156
AUTHORS	2 (bases 1 to 2960)
TITLE	Shimizu,N., Kitada,T. and Asakawa,S.
JOURNAL	Direct Submision
Submitted (22-DEC-1997)	Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
(E-mail:shimizu@mb.med.keio.ac.jp, Tel:81-3-3351-2370(ex.2720), Fax:81-3-3351-2370)	
FEATURES	Location/Qualifiers
source	1..2960
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="6"
	/map="6q25.2-q27"
	1..2960
gene	/gene="Parkin"
exon	1..108
	/gene="parkin"
	/number=1
	102..1499
CDS	/gene="parkin"
	/codon_start=1
	/product="Parkin"
	/protein_id="BAA25751.1"
	/db_xref="GI:3063388"
	/translation="MIVFVRENSSHGFPVEVDSTSIIFQLKEVAKRGVPADLRVYI FAGELKNDWTVQNCDDQOSIVHIVQRPWRGQEMNATGDDPNNAGCERPOSL TRVDLSSVLPBGDSGLAVILHTDSKQSPAPGSPAGRSIYNSFYVCKGCKOVQSG KLRYQSCSTCRQATILTTQGPSCWDVLIENRSGCOSPHCGTSABEFPFGCAHPSP DKREPVATRLIATNSRNITCITCTDVRSPVLYPOONSRYVLCICPHILYCTRLNDQ FVHDPOLYSLPCVACGSPSLKEILHFRFIIAGEBOYINRYQQYGAEECVLQGVLCPR PGCGAGLIPPEPDQRYTCRGGNGLGGFRPFCRCEAEVHEBCSAVFASTTTQOAR VDEAGAGLQAKWEAASKETLIKTTKPCPCRHVPEKNGCMHMKCPQPCRLEWCMNGC CEANRVCMGDHWDV"
exon	109..272
	/gene="parkin"
	/number=2
exon	273..513
	/gene="parkin"
	/number=3
exon	514..635
	/gene="parkin"
	/number=4
exon	636..719
	/gene="parkin"
	/number=5
exon	720..835
	/gene="parkin"
	/number=6
exon	836..972
	/gene="parkin"
	/number=7
exon	973..1034
	/gene="parkin"
	/number=8
exon	1035..1184

exon /gene="parkin"  
/number=9  
1185..1268  
/gene="parkin"  
/number=10  
1269..1386  
/gene="parkin"  
/number=11  
1387..2960  
/gene="parkin"  
/number=12  
exon

ORIGIN

Query Match 99.0%; Score 466.4; DB 9; Length 2960;  
Best Local Similarity 99.8%; Pred. No. 9,66-142;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGAAGTCCAGCAGGTAGATCAATCTACAAAGCTTTATGTATTTGCAAGGCCCTCTGT 60  
Db 504 GGAAGTCCAGCAGGTAGATCAATCTACAAAGCTTTATGTATTTGCAAGGCCCTCTGT 563

Qy 61 CAAAGATGTCAGCCCGGAAAACTCAGGGTACAGTCAGACCTGCAAGGCAAGCCTC 120  
Db 564 CAAAGATGTCAGCCCGGAAAACTCAGGGTACAGTCAGACCTGCAAGGCAAGCCTC 623

Qy 121 ACCTTGACCCAGGGGTCATCTTCTGGGATGATGTTTAAATTTCCAAACGGATGATGCT 180  
Db 624 ACCTTGACCCAGGGGTCATCTTCTGGGATGATGTTTAAATTTCCAAACGGATGATGCT 683

Qy 181 GAATGCCAATCCCAACATCTGCGCTGGAGCTAGTGCAGAAATTTCTTTAAATGTGAGCA 240  
Db 684 GAATGCCAATCCCAACATCTGCGCTGGAGCTAGTGCAGAAATTTCTTTAAATGTGAGCA 743

Qy 241 CACCCACCTCTGACAGAAACATCAGTACCTTGGACCTGATGCAACAAATAGTCGG 300  
Db 744 CACCCACCTCTGACAGAAACATCAGTACCTTGGACCTGATGCAACAAATAGTCGG 803

Qy 301 AACATCTTGCATTAAGTCGACAGACGTCAGAGGCCCGTCCTGGTTTCCAGTGCAC 360  
Db 804 AACATCTTGCATTAAGTCGACAGACGTCAGAGGCCCGTCCTGGTTTCCAGTGCAC 863

Qy 361 TCCCGCCACGTATTTGCTTGAAGCTGTTCCACTTATCTGTGTGACAAAGCTCATGAT 420  
Db 864 TCCCGCCACGTATTTGCTTGAAGCTGTTCCACTTATCTGTGTGACAAAGCTCATGAT 923

Qy 421 CGGAGATTGTTCAAGACCTTCAACTTGGCTAATCCCTGCTGTGTG 468  
Db 924 CGGAGATTGTTCAAGACCTTCAACTTGGCTAATCCCTGCTGTGTG 971

RESULT 8  
BC022014 1575 bp mRNA linear PRI 06-OCT-2003  
LOCUS Homo sapiens Parkinson disease (autosomal recessive, juvenile) 2,  
DEFINITION parkin, mRNA (cDNA clone MGC:26491 IMAGE:4824892), complete cde.  
ACCESSION BC022014  
VERSION BC022014.2 GI:34191069  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 1575)  
Strauberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Ditschenko,L., Maruina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Cavaan,T.L.,  
Schneer,T.E., Brownstein,M.J., Ueda,T.B., Tohbiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
22388257  
12477932  
2 (bases 1 to 1575)  
Strauberg,R.  
Direct Submission  
Submitted (22-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Aug 25, 2003 this sequence version replaced gi:18314632.  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
DNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshinuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbio.org](mailto:amadansystemsbio.org)  
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

FEATURES  
source  
1..1575  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:26491 IMAGE:4824892"  
/tissue\_type="Testis"  
/clone\_id="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

gene  
1..1575  
/gene="PARK2"  
/note="synonyms: AR-JP, PDI, PRKN"  
/db\_xref="locusid:5071"  
/db\_xref="WIM:602544"  
99..1262  
/gene="PARK2"  
/codon\_start=1  
/product="PARK2 protein"  
/protein\_id="AAH22014.1"  
/db\_xref="GI:18314633"  
/db\_xref="locusid:5071"  
/db\_xref="WIM:602544"  
/translation="MYPVPRNSHGPFVEVDSTSTFOLKEVYAKQGVADOLRV  
FAGELRNDWTVQNCDDQOSTVHVIVPWPKGGEWNAAGDDPRNNAAGCCERFOSL  
TRVDSVSLPGDSVGLAVILHTSRKDSPPAGSPARSINYSFVYCKGQVQPG  
KLKQCGTCRQATLTLTQGSCEMDVILPNMSEGCSPHCGTSAEFPGCAHPTS  
DKESVSLHLIATSRNITCTCTDVSFVLPQCSNRHVICLDFHLYCYTRLANDRO  
FVHDPQVGSILPCVGTGDTVTLRGALGSPRGVAGCPNSLTKELHHPRIILGEEQYNY  
QQYAEKCYLQMSVGLPRPGCAGLLEPDPQRVYTCGNGGLCGCGRTTK"  
99..317  
/gene="PARK2"  
/note="ubiquitin; Region: Ubiquitin family. This family  
contains a number of ubiquitin-like proteins: SUMO (smc3

ORIGIN  
homologue), Ned8, Elongin B, Rub1"  
/db\_xref="CDD:pfam00240"

Query Match 98.7%; Score 464.8; DB 9; Length 1575;  
Best Local Similarity 99.6%; Pred. No. 3e-141;  
Matches 466; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAAGTCAGAGGTAGATCAATCTACAAGCTTTTATGTATTCGAAGGCCCTCTG 60  
DB 501 GGAAGTCAGAGGTAGATCAATCTACAAGCTTTTATGTATTCGAAGGCCCTCTG 560  
QY 61 CAAAGAGTCAGCGCGGAAAACCTCAGGGTACAGTGCAGCACTGCAGGCAAGCGTC 120  
DB 561 CAAAGAGTCAGCGCGGAAAACCTCAGGGTACAGTGCAGCACTGCAGGCAAGCGTC 620  
QY 121 ACCCTGACCGAGGGTCCATCTTGTGGATGATGTTTAAATTCGAAACCGGATAGTGT 180  
DB 621 ACCCTGACCGAGGGTCCATCTTGTGGATGATGTTTAAATTCGAAACCGGATAGTGT 680  
QY 181 GAAATGCCAATCCCACTGCTGCGGAACTAGTGCAGAAATTTTCTTAAATGTGGAGA 240  
DB 681 GAAATGCCAATCCCACTGCTGCGGAACTAGTGCAGAAATTTTCTTAAATGTGGAGA 740  
QY 241 CACCCCACTCTGACAGAGAAATCATAGTCTTTCACCTGATGCAACCAATAGTCGG 300  
DB 741 CACCCCACTCTGACAGAGAAATCATAGTCTTTCACCTGATGCAACCAATAGTCGG 800  
QY 301 AACATCACTTGATTCGTGCAAGAGCGTCAAGAGCGCGTCTGTTTCCAGTGCAC 360  
DB 801 AACATCACTTGATTCGTGCAAGAGCGTCAAGAGCGCGTCTGTTTCCAGTGCAC 860  
QY 361 TCCCGCAGCGTATTTGCTTGAAGTCTTTCACCTTAACTGTGACAAGACTCAATGAT 420  
DB 861 TCCCGCAGCGTATTTGCTTGAAGTCTTTCACCTTAACTGTGACAAGACTCAATGAT 920  
QY 421 CGGCACTTGTTCAGAGCCCTCACTGTGCTACTCCCTGCTTGTGTG 468  
DB 921 CGGCACTTGTTCAGAGCCCTCACTGTGCTACTCCCTGCTTGTGTG 968

RESULT 9  
AF381282 1157 bp mRNA linear PRI 02-MAY-2002  
LOCUS AF381282  
DEFINITION Homo sapiens parkin isoform mRNA, complete cds, alternatively  
spliced.  
ACCESSION AF381282  
VERSION AF381282.1 GI:20385797  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1157)  
AUTHORS D'Agata,V., Scapagnini,G. and Cavallaro,S.  
JOURNAL Functional and molecular diversity of parkin  
UNPUBLISHED  
2 (bases 1 to 1157)  
AUTHORS D'Agata,V., Scapagnini,G. and Cavallaro,S.  
JOURNAL Direct Submission  
TITLE Submitted (01-MAY-2001) Institute of Bioimaging and Pathophysiology  
of the Central Nervous System, Italian National Research Council  
JOURNAL (CNR), Piazza Roma, 2, Catania 95123, Italy  
FEATURES  
SOURCE  
1..1157  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
333..1157  
/note="alternatively spliced; transcript variant 6"  
/codon\_start=1  
/product="parkin isoform"  
/protein\_id="AA021457.1"

/db\_xref="GI:20385798"  
/translation="MSGECQSPHCPGTSAEFFKCGAHPSTDKETPVALLIATNSRN  
ITCITCEVDRPVLVFOCNSRHVI CLDPCFI YCVTRINDROPVLDPOLGYSIPVAGC  
PNSLIKEIHFRIIGEDVNRXOQYAGAEFCVLCMGVLCRPGGABLLBEPDQKRT  
CEGNGGCGGAPFRCCKEKAHBEBCAVBASGTTTQARVDERAAQARWENASKE  
TIKRTYRCPKCHVPEKNGGCMKCPQPCRLIEWCMNCGEWNRCMDHWFV"

Query Match 97.8%; Score 460.8; DB 9; Length 1157;  
Best Local Similarity 99.6%; Pred. No. 5.9e-140;  
Matches 462; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GTCCAGCAGGTAGATCAATCTACAAGCTTTTATGTGTAATTCGAAGGCCCTGTCAA 64  
DB 166 GTCCAGCAGGTAGATCAATCTACAAGCTTTTATGTGTAATTCGAAGGCCCTGTCAA 225  
QY 65 GAGTGCAGCGGAAAACCTCAGGGTACAGTGCAGCACTGCAGGCAAGCGTCAACT 124  
DB 226 GAGTGCAGCGGAAAACCTCAGGGTACAGTGCAGCACTGCAGGCAAGCGTCAACT 285  
QY 125 TGACCCAGGGTCCATCTTGTGGATGATGTTTAAATTCGAAACCGGATAGTGTGAAT 184  
DB 286 TGACCCAGGGTCCATCTTGTGGATGATGTTTAAATTCGAAACCGGATAGTGTGAAT 345  
QY 185 GCCAATCCCACTGCGCTGGGACATAGTGCAGAAATTTTCTTAAATGTGGAGCAC 244  
DB 346 GCCAATCCCACTGCGCTGGGACATAGTGCAGAAATTTTCTTAAATGTGGAGCAC 405  
QY 245 CCACCTCTGACAGAGAAACATCATAGTCTTTCACCTGATGCAACCAATAGTCGAGACA 304  
DB 406 CCACCTCTGACAGAGAAACATCATAGTCTTTCACCTGATGCAACCAATAGTCGAGACA 465  
QY 305 TCACCTGATTAAGTGCAGAGACTGACAGAGCGCGTCTGTTTCCAGTGCACCTGCC 364  
DB 466 TCACCTGATTAAGTGCAGAGACTGACAGAGCGCGTCTGTTTCCAGTGCACCTGCC 525  
QY 365 GCCACGTGATTTGCTTGAAGTCTTTCACCTTAACTGTGACAAGACTCAATGATCGG 424  
DB 526 GCCACGTGATTTGCTTGAAGTCTTTCACCTTAACTGTGACAAGACTCAATGATCGG 585  
QY 425 AGTTTGTTCAGAGCCCTCACTGTGCTACTCCCTGCTTGTGTG 468  
DB 586 AGTTTGTTCAGAGCCCTCACTGTGCTACTCCCTGCTTGTGTG 629

RESULT 10  
AB060701 583 bp mRNA linear MAM 20-APR-2002  
LOCUS AB060701  
DEFINITION Bos taurus mRNA for parkin, partial cds.  
ACCESSION AB060701  
VERSION AB060701.1 GI:16444927  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1  
AUTHORS Kitada,T., Asakawa,S., Hattori,N., Matsumine,H., Yamamura,Y.,  
Mitsushima,S., Yokochi,M., Mizuno,Y. and Shimizu,N.  
TITLE Mutations in the parkin gene cause autosomal recessive juvenile  
parkinsonism  
JOURNAL Nature 392 (6676), 605-608 (1998)  
MEDLINE 98219084  
PUBMED 9560156  
REFERENCE 2  
AUTHORS Hasee,A., Yamada,H., Arai,K., Sunada,Y., Shimizu,T. and Matsumura,K.  
TITLE Characterization of parkin in bovine peripheral nerve  
JOURNAL Brain Res. 930 (1-2), 143-149 (2002)  
MEDLINE 21869006  
PUBMED 11879804  
REFERENCE 3 (bases 1 to 583)  
AUTHORS Hasee,A., Yamada,H., Arai,K., Shimizu,T. and Matsumura,K.

TITLE Direct Submission  
JOURNAL Submitted (25-APR-2001) Aseko Hase, Teikyo University School of Medicine, Dept. of Neurology and Neuroscience, 2-11-1, Kaga, Itabashi-ku, Tokyo 173-8605, Japan (E-mail:hase@med.teikyo-u.ac.jp, Tel:81-3-3964-1211(ex.1916), Fax:81-3-3964-6394)

FEATURES  
source  
1. 583  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="brain"  
/note="Identical sequence is also seen in peripheral nerve"

gene  
1. 583  
/gene="parkin gene"  
/db\_xref="GI:16444928"  
/protein\_id="BAB70670.1"  
/evidence="experimental"  
/product="parkin"  
/protein\_id="BAB70670.1"  
/translation="AGCEREPQSLTRVDSLSSVLPQDSVGLAVILHTSRKDSPPAGSPAGRSIYNSFFVYCKGRCQVOPGKLRVQDCTCRQATLTLTGSPSCMDVLIENRMS  
GPGSPHCPGTSAPFPKCGAHPDSKETSVALHLIATNSRNITCITDVRSPVLVF  
QNSRHVICLDCPHLYCVTRLNDRQFVHDPQLGY"

## ORIGIN

Query Match 96.4%; Score 454; DB 4; Length 583;  
Best Local Similarity 100.0%; Pred. No. 9.3e-138;  
Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGAAGTCCAGAGGTAGATCAATCTACAAAGCTTTATGATGATTCGAAAGGCCCTGT 60  
130 GGAAGTCCAGAGGTAGATCAATCTACAAAGCTTTATGATGATTCGAAAGGCCCTGT 189  
61 CAAAGATGACGAGCGGAAACTCAGAGGTACAGTGCAGCACTGCAGCAGCAAGCTC 120  
190 CAAAGATGACGAGCGGAAACTCAGAGGTACAGTGCAGCACTGCAGCAGCAAGCTC 249  
121 ACCTTGACCCAGGGTCCATCTTGGTGGATGATGATTTTAAATTCGAAACCGGATGATG 180  
250 ACCTTGACCCAGGGTCCATCTTGGTGGATGATGATTTTAAATTCGAAACCGGATGATG 309  
181 GAATGCCAATCCCACTGCTGCTGGAGTACTGAGCAATTTTCTTAAATGTGAGCA 240  
310 GAATGCCAATCCCACTGCTGCTGGAGTACTGAGCAATTTTCTTAAATGTGAGCA 369  
241 CACCCACCTCTGACAGGAACATCAGTACTTGACCTGATCGCAACAAATAGTCGG 300  
370 CACCCACCTCTGACAGGAACATCAGTACTTGACCTGATCGCAACAAATAGTCGG 429  
301 AACATCACTTGACATTAAGTGCAGAGGTGCAGAGCCCGGCTCTGTTTCCAGTGCAC 360  
430 AACATCACTTGACATTAAGTGCAGAGGTGCAGAGCCCGGCTCTGTTTCCAGTGCAC 489  
361 TCCCGCCAGCGATTTGCTTGAATGCTTTTCACTTAATGCTGACCAAGCTCAATGAT 420  
490 TCCCGCCAGCGATTTGCTTGAATGCTTTTCACTTAATGCTGACCAAGCTCAATGAT 549  
421 CCGCAGTTTGTTCAGACCCCTCAACTGGCTACT 454  
550 CCGCAGTTTGTTCAGACCCCTCAACTGGCTACT 583

RESULT 11  
AB039878 1564 bp mRNA linear ROD 16-APR-2004  
DEFINITION Rattus norvegicus mRNA for Parkin, complete cds.  
ACCESSION AB039878.1 GI:7229096  
VERSION Parkin.  
KEYWORDS Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

1 Hattori, N., Shimura, H., Kubo, S., Kitada, T., Wang, M., Asakawa, S., Minashima, S., Shimizu, N., Suzuki, T., Tanaka, K. and Mizuno, Y. Autosomal recessive juvenile parkinsonism: A key to understanding nigral degeneration in sporadic Parkinson's disease  
Neuropathology 20 81, 85-90 (2000)  
2 (bases 1 to 1564)  
Hattori, N., Wang, M. and Mizuno, Y.  
Direct Submission  
Submitted (07-MAR-2000) Nobutaka Hattori, Juntenodo University, Dept. of Neurology, Hongo 2-1-1, Bunkyo, Tokyo 113-8421, Japan (E-mail:hattori@med.juntendo.ac.jp, Tel:81-3-3813-3111(ex.3320), Fax:81-3-3813-7440)  
Location/Qualifiers

FEATURES  
source

1. 1564  
/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sraque-Dawley"

/db\_xref="taxon:10116"

/sex="male"

/clone\_id="recombinant bacterial type"

/tissue\_lib="adult"

/dev\_stage="adult"

/note="vector:pAP3neo"

1. 1564  
/gene="parkin"

/db\_xref="GI:7229097"

/translation="MIVFVRFNSSYGPEVDSPTISIFOLKEVAKRGYPADOLRVIFAGKELNHLTVQNCLEQOSIYHIVORPORKSHETNASGGDKFQSTPEGISIMPRSLTRYDLSHILPADSVGLAVILDPDSKDSRAARGRAKPTVHNSFFVYCKGCHVLOPGKLRYQCTGQATITTLAOGPSCMDVLIIPRMSECCSPDRCGRARFPKCAHPTS  
DKDTSVALNLITNNSRSIPCIACTDVNPVLPFCNHRHYICLDFHLYCVTRLNDRQFVHDAQGVSLPCVAGCPNSLIKELHNFRLIGEGVYRYYOYGAEECVLQWGVLCPRPGCGALLPREGORKVTCEGANGGCGFVFCRDKEXAHEGECDSMEBASGATQAYRVQRAARQARMEASKETIKKTKPCRCNVPIEKNGCMHKRPGQCKLWECMNG  
CEMRACMGHMPV"

## ORIGIN

Query Match 75.0%; Score 353.4; DB 10; Length 1564;  
Best Local Similarity 84.8%; Pred. No. 1.4e-104;  
Matches 396; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

2 GAATCCAGAGGTAGATCAATCTTACAAAGCTTTATGATGATTCGAAAGGCCCTGTC 61  
474 GAGGTCCAGAGGTAGATCAATCTTACAAAGCTTTATGATGATTCGAAAGGCCCTGTC 533  
62 AAAAGTGCAGCGCGGAAACTCAGGATACAGTGCAGACCTGCAGGAGGACAGCTCA 121  
534 ACAAGTGCAGCGCGGAAACTCAGGATACAGTGCAGACCTGCAGGAGGACAGCTCA 593  
122 CTTGACCCAGGGTCCATCTTGTCTGAGATGATGTTTAAATTCGAAACCGGATGATG 181  
594 CTTGACCCAGGGTCCATCTTGTCTGAGATGATGTTTAAATTCGAAACCGGATGATG 653  
182 AATGCCAATCCCACTGCTGCTGGAGTACTGAGCAAAATTTTCTTAAATGTGAGAC 241  
654 AGTGCATATCTCCAGACTGCTGCTGGAGTACTGAGCAAAATTTTCTTAAATGTGAGAC 713  
242 ACCCACTCTGACAAAGGAACATCAGTACTTGACCTGATGCGCAACAAATGTGAGCA 301  
714 ACCCACTCTGACAAAGGAACATCAGTACTTGATGATGATGATGATGATGATGATGAT 773  
302 ACATCACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361

Db 774 GCATCCCTGCATCGCGTGACGAGATGACAGAACCTGTCTTGCTTCCATGTAACC 833  
Qy 362 CCCGACAGTGAATTTGCTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 421  
Db 834 ACCGACAGTGAATTTGCTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 893  
Qy 422 GGCAGTTTGTTCAGACACCTCACTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 468  
Db 894 GGCAGTTTGTTCAGACACCTCACTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 940

RESULT 12  
AF343575 904 bp mRNA linear ROD 02-FEB-2002  
LOCUS Rattus norvegicus parkin transcript variant 7 mRNA, complete cds.  
DEFINITION AF343575  
ACCESSION AF343575.1 GI:18478869  
VERSION  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS 1 (bases 1 to 904)  
TITLE D'Agata,V.M., Scapagnini,G. and Cavallaro,S.  
JOURNAL Functional and molecular diversity of parkin in the rat brain  
REFERENCE 2 (bases 1 to 904)  
AUTHORS D'Agata,V.M., Scapagnini,G. and Cavallaro,S.  
TITLE Direct Submission  
JOURNAL Submitted (29-JAN-2001) Institute of Biomedicine and Pathophysiology of the Central Nervous System, Italian National Research Council (CNR), Piazza Roma, 2, Catania 95123, Italy  
Location/Qualifiers  
1..904  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
410..745  
/codon\_start=1  
/product="parkin transcript variant 7"  
/protein\_id="AAI7348.1"  
/db\_xref="GI:18478870"  
/translation="MSGCQSPDCPGRAPAEFFKGAHPISDKDTSVALNLTNNRS  
IPCIACIDVRNPVLVFCNHRHVICLDFCHLYCVTRINDRFVHDAQLGSLPCVVRF  
LPLGGGSP"

ORIGIN  
Query Match 74.4%; Score 350.2; DB 10; Length 904;  
Best Local Similarity 84.4%; Pred. No. 1.5e-103;  
Matches 394; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Db 2 GAAGTCAGAGTGAATGATCAATCAACACGCTTTTATGCTATGCAAGAGCCCTGTC 61  
Qy 240 GAGGTCCAGAGCTTAACCCACTACACAGCTTTTGTCTACTGCAAGAGCCCTGTC 299  
Db 62 AAAGATGACGCGGAAACTCAAGGTACAGTCAAGACCTGCAAGGAGGCAACGCTCA 121  
Qy 300 ACAAGTTCAGGCTGGGAACTCCGAGTTCACTGCGGACCTGCAAGGAGGCAACGCTCA 359  
Db 122 CCTTGACCCAGGCTGCTGCTGCGATGATGTTTAAATCCAAACCGATGATGCTG 181  
Qy 360 CTTTGCGCCAGGCGCCATCTTGCTGCGATGATGTTTAAATCCAAACCGATGATGCTG 419  
Db 182 AATGCCAATCCCACTGCGCTGCGGACTAGTGCAGATTTTCTTAAATGAGAGCAC 241  
Qy 420 AGTGCCATCTCCAGACTGCGCTGCGGACAGAGCTGAATTTTCTTAAATGAGAGCAC 479  
Db 242 ACCCCACTTGTGACAGAGAAACATCACTGATGCTTTGCACTGATGCAACAAATAGTGG 301  
Qy 480 ACCCAACTGACAGAGAAACATCACTGATGCTTTGCACTGATGCAACAAAGGCGGCA 539  
Db 302 ACATCACTTGCATTAAGTGCAGAGAGCTCAAGAGCCCGGCTGCTTCCATGCACT 361

Db 540 GCATCCCTGCATCGCGTGACGAGATGACAGAACCTGTCTTGCTTCCATGTAACC 599  
Qy 362 CCCGACAGTGAATTTGCTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 421  
Db 600 ACCGACAGTGAATTTGCTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 659  
Qy 422 GGCAGTTTGTTCAGACACCTCACTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 468  
Db 660 GGCAGTTTGTTCAGACACCTCACTTGAATGTTTGCATTAATGCTGTGACAAAGCTCAATGATAC 706

RESULT 13  
AF168004 1466 bp mRNA linear ROD 19-FEB-2000  
LOCUS Rattus norvegicus parkin mRNA, partial cds.  
DEFINITION AF168004  
ACCESSION AF168004.1 GI:7001383  
VERSION  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS 1 (bases 1 to 1466)  
TITLE D'Agata,V., Zhao,W. and Cavallaro,S.  
JOURNAL Cloning and distribution of the rat parkin mRNA  
REFERENCE 2 (bases 1 to 1466)  
AUTHORS D'Agata,V. and Cavallaro,S.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUL-1999) Institute of Biomedicine and Pathophysiology of the Central Nervous System, Italian National Research Council, Piazza Roma, 2, Catania 95123, Italy  
Location/Qualifiers  
1..1466  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/tissue\_type="brain"  
91..2146  
/codon\_start=1  
/product="Parkin"  
/protein\_id="AAI34874.1"  
/db\_xref="GI:7001384"  
/translation="MIVFVRFNSYGFVPEVDSPTSICQLKEVAKRGQVPAQQLRVI  
FAGKEIQLNHLTVQNCIDLEOOSIVHIVORPORSKSHETNAGSDKPOSTPEGSIMPEPSL  
TRVDLSHILPADSVGLAVILDTDSKSDSEARGBAKPTYSFVYCKGPKHQPQ  
KLRVCGTCROATVTLAOGSPSCMDVLIENRMSGCQSPDCGTAABFFKGAHPIS  
DKDTSVALNLTNNRSITCICTGATVRNPVLVFCNHRHVICLDFCHLYCVTRINDRO  
FVHDAQLGSLPCVVRVACGNSLKEIHLHFRILDEBQINRYQYGAEECVLQMGVLCPR  
PGGAGLPEQGVKVTGCGGNGLGCGFPCDCKEAHEGSDIMEFSGATISQAYR  
VDORAEQARMEBAESKEITIKTKTKPCRCNPVIEKNKGCMHMKCPQCKLEMCWNG  
CEMNACMG"

ORIGIN  
Query Match 74.0%; Score 348.6; DB 10; Length 1466;  
Best Local Similarity 84.2%; Pred. No. 5.3e-103;  
Matches 353; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Db 2 GAAGTCAGAGTGAATGATCAATCAACACGCTTTTATGCTATGCAAGAGCCCTGTC 61  
Qy 494 GAGGTCCAGAGCTTAACCCACTACACAGCTTTTGTCTACTGCAAGAGCCCTGTC 553  
Db 62 AAAGATGACGCGGAAACTCAAGGTACAGTCAAGACCTGCAAGGAGGCAACGCTCA 121  
Qy 554 ACAAGTTCAGGCTGGGAACTCCGAGTTCACTGCGGACCTGCAAGGAGGCAACGCTCA 613  
Db 122 CCTTGACCCAGGCTGCTGCTGCGATGATGTTTAAATCCAAACCGATGATGCTG 181  
Qy 614 CTTTGCGCCAGGCGCCATCTTGCTGCGATGATGTTTAAATCCAAACCGATGATGCTG 673

Qy		182	AATGCGAATCCCAACACTGCCTGGAGATGTATGCAAGAATTTTCTTTAAATGTGAAGAC	244
Db		674	AGTGCCAATCTCCAGACTGCCCTGGACAAGACGTAATTTTTCTTTAAATGTGAAGAC	733
Qy		242	ACCCCACTCTGACAAAGAAAATCATCAGTAGCTTTTGACCTGTATGCAACAATAATGTGGA	301
Db		734	ACCCAACTCAGACAAAGAACATCATGAGCTTTGAACTGTATCACCAACAACAGCCGCA	793
Qy		302	ACATCACTTGGATTACGTGCAACAGCTCAGAGAGCCCGCTCTGGTTTTCAGTGCACACT	361
Db		794	GCAATCCCCCTGATGCGCTGCACAGGATGCAGAAACCCCTGTCTTCCAAATGTAAAC	853
Qy		362	CCCGCAACGTATTGCTTAGACTGTTCCTTCCACTTATATCTGTGTGCAAGACTCATGATC	421
Db		854	ACCCGCACTGATCTGTTTGGACTGCTTCCACTTGTACTGTGTACANAGCTCAACATC	913
Qy		422	GGCAGTTTGTTCACGACACCCCTCAACTGGCTACTCCCTGCTGTGTG	468
Db		914	GGCAGTTTGTCCACGACGCTCAGCTTGCTACTGCTCGCTCGTGTG	960
RESULT	14			
AF210434				
LOCUS		1576 bp	mRNA	linear ROD 02-JUL-2002
DEFINITION		Rattus norvegicus parkin mRNA, complete cds.		
ACCESSION		AF210434		
VERSION		AF210434.1 GI:11527823		
KEYWORDS				
SOURCE		Rattus norvegicus (Norway rat)		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE		1 (bases 1 to 1576)		
AUTHORS		Gu,W.-J., Hirsch,E.N., Lagunes,M.Z., Parent,A., Pradier,L., Bohme,G.A., Agid,Y., Hirsch,E.C., Ralsman-Vozari,I., and Brice,A.		
JOURNAL		Cloning of rat parkin cDNA and distribution of parkin in rat brain		
MEDLINE		J. Neurochem. 74 (4), 1773-1776 (2000)		
PUBMED		20199898		
REFERENCE		2 (bases 1 to 1576)		
AUTHORS		Gu,W.-J. and Abbas,N.E.		
TITLE		Direct Submision		
JOURNAL		Submitted (02-DEC-1999) Neurology, INSERM U289, 47, Boulevard de l'Hopital, Paris 75013, France		
FEATURES		Location/Qualifiers		
Source		1..1576		
		/organism="Rattus norvegicus"		
		/mol_type="mRNA"		
		/db_xref="taxon:10116"		
		/feature_type="hypothalamus"		
		63..1460		
		/codon_start=1		
		/product="parkin"		
		/protein_id="AAG37013.1"		
		/db_xref="GI:11527824"		
		/translation="MIVFVRNFSYGFPEVVDSTSI.FQLKEVVAKRGVPADLRVLYPAGELIENHLTVONKCDLEQSLIVITVORPKSKHETNAASGXDSPTDEGSIMPRSLSTRVLDSSHTLPADSVGLAVITDPTDKSSSEARBPBAPKTTHSFVTCKGKCHVQPREKLRVQCGCTGRQATLTLLAQGPSICWDVLIPNNMSBCSPDPGTRAEFFFGCAHPGSRKDSDVALMLITNNRSRIPICTACTDVNRPLVLFQCNHRHYCLDCFHLYCVTRLNDQRKFVHDALQGLSLPFCVAGCPNSLIKEILHFRILIGEQYNRYOQYGABECLQMGVLCRRPGCGAGLIPBEQGOKKVKTECBGNGAGCGVFPCRDCKEAYHEGECDSMFASGATSQAVRKYVDQAABQABREASEASKRTIKTKTTPRCPCNVPIEKNGSCMHMKPPQPKLKEMCMNCGCERNRACMGDMWPDV"		
ORIGIN				
Query Match		74.0%	Score 348.6;	DB 10; Length 1576;
Best Local Similarity		84.2%;	Pred. No. 5-4e-103;	
Matches	393;	Conservative	0;	Mismatches 74; Indels 0; Gaps 0;
Qy	2	GAAGTCCACGACGGTGTGATCAATCTTTCACACAGCTTTTATGTGTATTGCCAAGGCCCTCTC	61	

Db	Query Match	Best Local Similarity	73.9%	Score 348.2	DB 10	Length 1157
Db	466	GAGSTCCAGAAAGCTAAACCACCTACACAGCTTTTGTCTACTGCAAAAGCCCTGGCC	84.3%	Pred. No. 7e-103		525
Qy	62	AAAGATGTCAGCCGGGAAAACTCAGGGTACAGTCACGACGACTGCAAGCGCAACGCTCA				121
Db	526	ACAAGTCCAGGACCTGGGAAACTCCGAGTTCAGTGGGCACTTCGACGACCAAGCAACCTCA				585
Qy	122	CCTTGACCCGAGGGTCCATCTTTGCTGGGATGATGTTTAAATTCGAAACCGATGAGTGCTG				181
Db	586	CTTTGGGCGAGGGGCCATCTTGTCTGGGATGATGTTTAAATTCGAAACCGATGAGTGAGTG				645
Qy	182	AATGCAATCCCCACATCGCCCTGGGACTAGTGAGAAATTTTCTTTAAATGTGAGGAC				241
Db	646	AGTCCAAATCTCCAGATCGCCCTGGGACCAAGAGCTGAATTTTCTTTAAATGTGAGGAC				705
Qy	242	ACCCACCTCTGACAGAAACATCACTAGTTCGACTGATGCAACCAAAATAGTCGA				301
Db	706	ACCCAACTCAGACAAAGACACATCACTAGTTCGACTGATGCAACCAAGAGTCGA				765
Qy	302	ACATCATCTTGCTATTAGTGCACAGAGCTCAGAGACCCCTGCTCTGTTTCCAGTGCACCT				361
Db	766	GCAATCCCTGCATCGCTGCACGAGATGCAGGAACCTGTCTTGGTCTTCCAAATGTAAC				825
Qy	362	CCGGCCACGATTTGGTTGTTAGACTGTTTCCACTATCTGTGTGACAAAGACTCAATGATC				421
Db	826	ACCCACGATGATGTTTGGACTGCTTCCACTTGTACTGTGTACAAAGGCTCAACGATC				885
Qy	422	GGCAGTTTGTTCACGACCCCTCAACCTTGCTGCTACTCCCTGCTTGTGTG				468
Db	886	GGCAGTTTGTTCACGACGCTCAGCTGCTGCTACTGCTGCTGCTGTGTG				932
RESULT 15	AF343574	1157 bp	mRNA	linear	ROD 02-FEB-2002	
LOCUS	AF343574					
DEFINITION	Rattus norvegicus parkin transcript variant 6 mRNA, complete cds.					
ACCESSION	AF343574					
VERSION	AF343574.1	GI:18478865				
KEYWORDS						
ORGANISM	Rattus norvegicus (Norway rat)					
SOURCE	Rattus norvegicus					
REFERENCE	1 (bases 1 to 1157)					
AUTHORS	D'Agata,V.M., Scapagnini,G. and Cavallaro,S.					
TITLE	Functional and molecular diversity of parkin in the rat brain					
JOURNAL	Unpublished					
AUTHORS	2 (bases 1 to 1157)					
TITLE	D'Agata,V.M., Scapagnini,G. and Cavallaro,S.					
JOURNAL	Submitted (29-JUN-2001) Institute of Biomedicine and Pathophysiology of the Central Nervous System, Italian National Research Council (CNR), Piazza Roma, 2, Catania 95123, Italy					
FEATURES	Location/Qualifiers					
source	1..1157					
	/organism="Rattus norvegicus"					
	/mol_type="mRNA"					
	/db_xref="taxon:10116"					
	333..1157					
	/codon_start=1					
	/product="parkin transcript variant 6"					
	/protein_id="AA173348.1"					
	/db_xref="GI:18478865"					
	/translation="MSGECSPDCPGTAEFFFKGAPTSDDKTSVALNI,ITNNSRS					
	IPICACIDVNPVLPVPCNHRHYICLDCPLCYCTRILNDROPVADQLGSLPCVAGC					
	PENGIKRLHFRILGEEQYNRKYQYGAEEVYLQMGVLCRPGCGAALBPQCKYTT					
	CSEGGNGGCGFVCRDCKEAYHSEGCSEWSEBASGATQARVDDRAAEQAWESASVE					
	TIKTTYPCRCNVPIEKNGCMHMKPOQCKLEWCMNGCEWNRACMGDHWEDV"					
ORIGIN						





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 25, 2005, 15:46:49 ; Search time 265.776 Seconds  
(without alignments)  
10490.795 Million cell updates/sec

Title: US-10-622-817-10

Perfect score: 471

Sequence: 1 ggaagtcacagcagtagatc.....accctcgtctgtgtcag 471

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	471	100.0	471	5	AAH77664
2	468	99.4	2960	3	AAA46765
3	466.4	99.0	2960	2	AAK99923
4	466.4	99.0	2960	8	AAD47679
5	466.4	99.0	2960	8	AAD47680
6	466.4	99.0	2960	8	ACD13332
7	340.2	72.2	3092	4	AAFS5253
8	340.2	72.2	3092	4	AAFS5257
9	340.2	72.2	3253	4	AAFS5256
10	340.2	72.2	3253	4	AAFS5254
11	340.2	72.2	3254	4	AAFS5255
12	340.2	72.2	3255	4	AAFS5259
13	340.2	72.2	3255	4	AAFS5244
14	340.2	72.2	3255	4	AAFS5260
15	340.2	72.2	3255	4	AAFS5258
16	338.6	71.9	3014	2	AAFS5247
17	288.4	61.2	2876	2	AAK99924
18	288.4	61.2	2876	8	ACD13333
19	265.4	56.3	3136	4	AAFS5250
20	262.8	55.8	2895	4	AAFS5248

21	251.8	53.5	2513	8	ACD13334	ACD13334 Human DNA
22	240.8	51.1	857	4	AAFS5246	AAFS5246 Nucleoid
23	239.8	50.9	1459	4	AAFS5245	AAFS5245 Nucleoid
24	198.8	42.2	3170	4	AAFS5251	AAFS5251 Nucleoid
25	100.4	21.3	318	12	ADM96486	ADM96486 Modified
26	90.4	19.2	321	12	ADM96488	ADM96488 Modified
27	78.8	16.7	2918	4	AAFS5252	AAFS5252 Nucleoid
28	74	15.7	518	12	ADM96490	ADM96490 Murine pa
29	71	15.1	1458	4	ABL16595	ABL16595 Drosophila
30	53	11.3	470	12	ADM96491	ADM96491 Murine pa
31	50.2	10.7	65	6	ABN3295	ABN3295 Mouse api
32	46.4	9.9	3865	4	ABL16594	ABL16594 Drosophila
33	35.6	7.6	1449	10	ADE07643	ADE07643 Novel cod
34	34.8	7.4	361	3	AAFC5989	AAFC5989 Human lun
35	34.8	7.4	361	6	ABL49208	ABL49208 Human lun
36	34.8	7.4	361	6	ABO92394	ABO92394 Human lun
37	34.8	7.4	361	9	ADA28383	ADA28383 Human lun
38	34.8	7.4	361	10	ADH36947	ADH36947 Human lun
39	34.8	7.4	361	12	ADM56750	ADM56750 Human sec
40	34.8	7.4	405	3	AAFC07249	AAFC07249 Human sec
41	33.6	7.1	362	10	ADFC0743	ADFC0743 Leukaemia
42	33.6	7.1	395	6	ABO56883	ABO56883 Human col
43	33.6	7.1	446	12	ADO40914	ADO40914 Human CDN
44	33.6	7.1	447	9	ACH26909	ACH26909 Human adu
45	33.6	7.1	483	9	ACH26764	ACH26764 Human adu

## ALIGNMENTS

RESULT 1  
AAH77664  
ID AAH77664 standard; DNA; 471 bp.  
XX  
AC AAH77664;  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Nucleotide sequence of human Parkin protein fragment.  
XX  
KW Human; Parkin-Associated Protein 1; PAPI; Parkin gene;  
KW neurodegenerative disease; Parkinson's disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200160857-A2.  
XX  
PN 23-AUG-2001.  
XX  
PD 15-FEB-2001; 2001WO-FR00461.  
XX  
PF 17-FEB-2000; 2000FR-00001980.  
XX  
PR 18-APR-2000; 2000US-0198489P.  
PA (AVET ) AVENTIS PHARMA SA.  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PI Koutnikova H, Brice A, Fournier A, Pradier L, Prades C;  
PI Arnould-Reguigne I, Rosier-Montus M, Corti O;  
PI WPI; 2001-550047/61.  
DR P-PSDB; AAG67213.  
XX  
PT A new protein, designated Parkin-Associated Protein 1 (PAPI), is an  
PT interaction partner of Parkin and is useful to treat neurodegenerative  
PT pathologies including Parkinson's disease.  
XX  
PS Example 1; Page 60-61; 82pp; French.  
XX  
CC The present sequence encodes the central region of a human Parkin  
CC protein. The protein was used to identify Parkin-Associated Protein 1  
CC (PAPI) protein. PAPI is associated with the Parkin gene, which is mutated  
CC in certain forms of familial (juvenile autosomal recessive) Parkinson's

CC disease. PAPI has some homology with synaptotagmins. PAPI is used to  
 CC treat neurodegenerative diseases, particularly to diagnose and treat  
 CC Parkinson's disease  
 XX  
 SQ Sequence 471 BP; 118 A; 127 C; 109 G; 117 T; 0 U; 0 Other;

Query Match 100.0%; Score 471; DB 5; Length 471;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-153;  
 Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGTCCAGCAGGTAGATCAATCTAACAACACCTTTTATGTATGCAAAAGCCCTGT 60  
 Db 1 GGAAGTCCAGCAGGTAGATCAATCTAACAACACCTTTTATGTATGCAAAAGCCCTGT 60  
 QY 61 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGAGGCAACGCTC 120  
 Db 61 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGAGGCAACGCTC 120  
 QY 121 ACCTTGACCCAGGGTTCATCTTGTGGATGATGTTTAAATTCAAACCGGATGAGTGT 180  
 Db 121 ACCTTGACCCAGGGTTCATCTTGTGGATGATGTTTAAATTCAAACCGGATGAGTGT 180  
 QY 181 GAATGCCAATCCCACTGACCTGGGACTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
 Db 181 GAATGCCAATCCCACTGACCTGGGACTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
 QY 241 CACCCACCTCTGACAGAAACATCAGTAGCTTGCACCTGATGCACAAATATGTCG 300  
 Db 241 CACCCACCTCTGACAGAAACATCAGTAGCTTGCACCTGATGCACAAATATGTCG 300  
 QY 301 AACATCACTTGATTAAGTGCAGACAGTCAAGAGCCCGTCTGTTTCCAGTGCAC 360  
 Db 301 AACATCACTTGATTAAGTGCAGACAGTCAAGAGCCCGTCTGTTTCCAGTGCAC 360  
 QY 361 TCCCGCAGTATTTGCTTAGACTGTTTCCATTACTGTGTGACAAAGCTCAATGAT 420  
 Db 361 TCCCGCAGTATTTGCTTAGACTGTTTCCATTACTGTGTGACAAAGCTCAATGAT 420  
 QY 421 CGGCACTTTGTTCAAGACCTTCAACTTGGCTACTCCCTGCTGTGTGTAG 471  
 Db 421 CGGCACTTTGTTCAAGACCTTCAACTTGGCTACTCCCTGCTGTGTGTAG 471

## RESULT 2

AAA46765  
 ID AAA46765 standard; cDNA; 2960 BP.

AAA46765;

25-SEP-2000 (first entry)

CDNA sequence of the human parkin protein gene.

Human; parkin protein; parkin gene; Parkinson's disease;  
 anti-Parkinson agent; ss.

Homo sapiens.

W0200031253-A2.

02-JUN-2000.

18-NOV-1999; 99WO-FR002833.

19-NOV-1998; 98FR-00014524.

12-MAR-1999; 99US-0124239P.

04-AUG-1999; 99FR-00010140.

(RHON ) RHONE-POULENC ROBER SA.

(INRM ) INST NAT SANTE & RECH MEDICALE.

Brice A, Lucking C, Abbas NE, Denefle P, Ricard S, Bouley S;

DR WPI, 2000-411952/35.

XX New variant forms of the human parkin gene, used as source of primers and  
 PT probes for detecting susceptibility to Parkinson's disease.

XX Claim 24; Fig 1; 71pp; French.

PS The present sequence represents a cDNA sequence of the human parkin  
 CC protein gene. The specification describes a parkin gene which has genetic  
 CC alterations. Cells, or transgenic animals, that express the altered  
 CC parkin gene are used to screen for compounds that can counter the effects  
 CC of a genetic alteration in the parkin gene, or more generally for  
 CC studying the properties of the parkin protein. Detection of the specified  
 CC alterations is used to diagnose susceptibility to Parkinson's disease.  
 CC The modified polynucleotide is also used to express the corresponding  
 CC protein, which is then used to screen for potential anti-Parkinson agents  
 CC and to raise antibodies (for detecting variants of parkin protein)

XX Sequence 2960 BP; 815 A; 726 C; 698 G; 721 T; 0 U; 0 Other;

Query Match 99.4%; Score 468; DB 3; Length 2960;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-151;  
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGTCCAGCAGGTAGATCAATCTAACAACGCTTTATGTATGCAAAAGCCCTGT 60  
 Db 504 GGAAGTCCAGCAGGTAGATCAATCTAACAACGCTTTATGTATGCAAAAGCCCTGT 563  
 QY 61 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGAGGCAACGCTC 120  
 Db 564 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGAGGCAACGCTC 623  
 QY 121 ACCTTGACCCAGGGTTCATCTTGTGGATGATGTTTAAATTCAAACCGGATGAGTGT 180  
 Db 121 ACCTTGACCCAGGGTTCATCTTGTGGATGATGTTTAAATTCAAACCGGATGAGTGT 180  
 QY 181 GAATGCCAATCCCACTGACCTGGGACTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
 Db 181 GAATGCCAATCCCACTGACCTGGGACTAGTGCAGAAATTTTCTTAAATGTGAGCA 240  
 QY 241 CACCCACCTCTGACAGAAACATCAGTAGCTTGCACCTGATGCACAAATATGTCG 300  
 Db 241 CACCCACCTCTGACAGAAACATCAGTAGCTTGCACCTGATGCACAAATATGTCG 300  
 QY 301 AACATCACTTGATTAAGTGCAGACAGTCAAGAGCCCGTCTGTTTCCAGTGCAC 360  
 Db 301 AACATCACTTGATTAAGTGCAGACAGTCAAGAGCCCGTCTGTTTCCAGTGCAC 360  
 QY 361 TCCCGCAGTATTTGCTTAGACTGTTTCCATTACTGTGTGACAAAGCTCAATGAT 420  
 Db 361 TCCCGCAGTATTTGCTTAGACTGTTTCCATTACTGTGTGACAAAGCTCAATGAT 420  
 QY 421 CGGCACTTTGTTCAAGACCTTCAACTTGGCTACTCCCTGCTGTGTGTAG 468  
 Db 924 CGGCACTTTGTTCAAGACCTTCAACTTGGCTACTCCCTGCTGTGTGTAG 971

## RESULT 3

AAK99923  
 ID AAK99923 standard; DNA; 2960 BP.

AAK99923;

21-OCT-1999 (first entry)

Human parkin gene variant.

Parkinson's disease related gene; parkin gene; variant; gene therapy; ss.

Homo sapiens.

Key location/Qualifiers  
 CDS 102..1499

```
PT XX /*tag= a
XX PN MO9940191-A1.
XX PD 12-AUG-1999.
XX PF 09-FEB-1999; 99MO-JP000545.
XX PR 09-FEB-1998; 98JP-00027531.
XX PA (SHIM/) SHIMIZU N.
XX PA (MIZU/) MIZUNO Y.
XX PI Shmizu N, Mizuno Y;
XX DR WPI; 1999-494295/41.
XX DR P-PSDB; AAY32501.
XX PT Gene implicated in the pathology of Parkinson's disease, used for
XX PT treatment of the disease.
XX PS Claim 1, Page 83-88; 114p; English.
XX CC This sequence represents a gene of the invention, and is implicated in
XX CC the pathology of Parkinson's disease. This sequence is a variant of the
XX CC parkin gene found in Parkinson's disease patients. The sequences may be
XX CC used for the diagnosis, treatment (including gene therapy) and
XX CC investigation of Parkinson's disease
XX SQ Sequence 2960 BP; 815 A; 727 C; 698 G; 720 T; 0 U; 0 Other;

Query Match 99.0%; Score 466.4; DB 2; Length 2960;
Best Local Similarity 99.8%; Pred. No. 8.4e-151;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCACAGAGTATCATCTACACAGCTTTTATGTTGTAATGCAAGGCCCTGT 60
DB 504 GGAAGTCACAGAGTATCATCTACACAGCTTTTATGTTGTAATGCAAGGCCCTGT 563
QY 61 CAAAGATGACAGCCGGGAAAACTCAGGGTACAGTCGACGACCTTCAGAGGCAACGCTC 120
DB 564 CAAAGATGACAGCCGGGAAAACTCAGGGTACAGTCGACGACCTTCAGAGGCAACGCTC 623
QY 121 ACCTTGACCAAGGGTCCATCTTGCTGGGATGATGTTTAAATTCGAAACCGGATGATG 180
DB 624 ACCTTGACCAAGGGTCCATCTTGCTGGGATGATGTTTAAATTCGAAACCGGATGATG 683
QY 181 GAATGCCAATCCCAACAGTCCCTGGGACTAGTCAGAAATTTTCTTAAATGTGAGCA 240
DB 684 GAATGCCAATCCCAACAGTCCCTGGGACTAGTCAGAAATTTTCTTAAATGTGAGCA 743
QY 241 CACCCACCTCTGACAGAAAGAAACATGATGATCTTGACCTGATGCAACAAATAGTCG 300
DB 744 CACCCACCTCTGACAGAAAGAAACATGATGATCTTGACCTGATGCAACAAATAGTCG 803
QY 301 AACATCACTTGACATTAAGTCAGAGAGTCCGAGGCCCCCTCGTTCCTGATGCAAC 360
DB 804 AACATCACTTGACATTAAGTCAGAGAGTCCGAGGCCCCCTCGTTCCTGATGCAAC 863
QY 361 TCCGCGACAGTATTTGCTTGAAGTCTGTTTCCATTATAGTGTGACAAAGCAATGAT 420
DB 864 TCCGCGACAGTATTTGCTTGAAGTCTGTTTCCATTATAGTGTGACAAAGCAATGAT 923
QY 421 CGGAGTTTGTTCAGACACCTCAACTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 468
DB 924 CGGAGTTTGTTCAGACACCTCAACTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 971

RESULT 4
AAD47679 standard; cDNA; 2960 BP.
XX AAD47679
XX AAD47679;
```

```
XX XX 24-FEB-2003 (first entry)
XX DT Human Parkin protein encoding cDNA.
XX DE
XX KW Human; Parkin protein; neurological disorder; apoptosis; gene therapy;
XX KW ischaemic stroke; Parkinson's disease; Alzheimer's disease; noctropic;
XX KW transgenic; cerebroprotective; neuroprotective; neurotransplantation;
XX KW gene; 86.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 102..1499
XX FT /*tag= a
XX FT /product= "Human Parkin protein"
XX PN MO200279459-A2.
XX PD 10-OCT-2002.
XX PP 02-APR-2002; 2002MO-DK000221.
XX PR 29-MAR-2001; 2001DK-00000525.
XX PR 03-APR-2001; 2001US-0281286P.
XX PA (NSGE-) NSGENE AS.
XX PI Jensen PH;
XX PI
XX DR WPI; 2003-046812/04.
XX DR P-PSDB; AAE30800.
XX FT New isolated nucleic acid sequence encoding a Parkin polypeptide, useful
XX FT for treating, preventing or diagnosing neurological disorders, e.g.
XX FT Parkinson's disease, Alzheimer's disease or ischemic stroke, and in
XX FT screening assays.
XX PS Example 1, Page 68-69; 71p; English.
XX CC The invention relates to Parkin protein and its corresponding nucleic
XX CC acid sequence. The nucleic acid sequence is useful for altering the
XX CC proteolytic processing of Parkin at its potential cleavage site at Asp
XX CC 126. The invention is used in manufacturing or testing a pharmaceutical
XX CC composition for treating and/or preventing a neurological disorder, e.g.
XX CC Alzheimer's disease or ischaemic stroke. It also used for detecting the
XX CC occurrence of proteolytic processing of Parkin at Asp 126 in a sample, in
XX CC monitoring a potential disposition for a neurodegenerative disease, and
XX CC for treating, preventing and/or diagnosing Parkinson's disease or other
XX CC neurodegenerative disorders. The viral vector is used for transforming
XX CC neuronal cells in vivo or ex vivo. The invention is useful for
XX CC neurotransplantation into the CNS of a mammal. It may be used in
XX CC screening assays to identify compounds that increase or decrease
XX CC apoptosis. It is also used in gene therapy. The present sequence is human
XX CC Parkin protein encoding cDNA
XX SQ Sequence 2960 BP; 815 A; 727 C; 698 G; 720 T; 0 U; 0 Other;

Query Match 99.0%; Score 466.4; DB 8; Length 2960;
Best Local Similarity 99.8%; Pred. No. 8.4e-151;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCACAGAGTATCATCTACACAGCTTTTATGTTGTAATGCAAGGCCCTGT 60
DB 504 GGAAGTCACAGAGTATCATCTACACAGCTTTTATGTTGTAATGCAAGGCCCTGT 563
QY 61 CAAAGATGACAGCCGGGAAAACTCAGGGTACAGTCGACGACCTTCAGAGGCAACGCTC 120
DB 564 CAAAGATGACAGCCGGGAAAACTCAGGGTACAGTCGACGACCTTCAGAGGCAACGCTC 623
QY 121 ACCTTGACCAAGGGTCCATCTTGCTGGGATGATGTTTAAATTCGAAACCGGATGATG 180
DB 624 ACCTTGACCAAGGGTCCATCTTGCTGGGATGATGTTTAAATTCGAAACCGGATGATG 683
```





SQ Sequence 3043 BP; 806 A; 763 C; 737 G; 737 T; 0 U; 0 Other;  
Query Match 72.2%; Score 340.2; DB 4; Length 3043;  
Best Local Similarity 84.0%; Pred. No. 7e-107;  
Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 11 CAGGTAGATCAATCTACAAAGCTTTATGTATGATGCAAGAGCCCTGTCAAGAGTGC 70  
DB 538 CAGTTAAACCACTCAACAGCTTTTATCTACAGCAAGAGCCCTGTCAAGAGTGC 597  
QY 71 ACCCGGAAAACCTCAGAGGTCAGTCAGACCTGCGAGGAGCAAGCTCAGCTTGACCC 130  
DB 598 AGCTTGGAAAGCTCCGAGTTCAGTGTGCGACCTGCAAAACCAACCTCAGCTTGAGCC 657  
QY 131 AGGTCATCTTGTCTGGATGATGTTTAAATTCACAAACCGATGAGTGTGAATGCCAAT 190  
DB 658 AGGCGCATCTTGTCTGGACGATGTTTAAATTCACAAACCGATGAGTGTGAATGCCAAT 190  
QY 191 CCCCACTGCTGCTGGAGTGTGACAAATTTTCTTAAATGTGAGACACCCCACT 250  
DB 718 CTCAGACTGCTGGAACCAAGCTGAATTTTCTTAAATGTGAGACACCCCACT 777  
QY 251 CTGACAAAGAAACATCAGTACGTTTGCACCTGATGCAACAAATGTGGAACATCACTT 310  
DB 778 CAGACAAAGAACGTCGTGATGTTGAACCTGATCACAAGCAAGGCGCAGCATCCCTT 837  
QY 311 GCATTACGTGACAGACGTCAGAGCCCGTCTGTTTCCAGTGCACACTCCGCGCAG 370  
DB 838 GCATTACGTGACAGACGTCAGAGCCCGTCTGTTTCCAGTGCACACTCCGCGCAG 370  
QY 371 TGATTGCTTGAAGCTTTTCCATTATCTGTGTGACAAAGCTCAATGATGGCAGTTTG 430  
DB 898 TGATCTGTTTGAAGCTTTTCCATTATCTGTGTGACAAAGCTCAATGATGGCAGTTTG 430  
QY 431 TTCACGACCCCTCAATGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467  
DB 958 TCCACGATGCTCAATGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 994  
RESULT 8  
AAFS5257  
ID AAF55257 standard; cDNA, 3092 BP.  
AC AAF55257;  
XX  
DT 29-MAY-2001 (first entry)  
XX  
DE Nucleotide sequence of murine parkin2 with a frameshift mutation.  
XX  
KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
KW brain tumour; head trauma; stroke; vascular irregularity;  
KW metabolic irregularity; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 129..173  
FT /tag= a  
FT /product= "truncated parkin2"  
XX  
EN EP1081225-A1.  
XX  
PD 07-MAR-2001.  
XX  
PF 30-AUG-1999; 99EP-00116766.  
XX  
PR 30-AUG-1999; 99EP-00116766.  
XX  
PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
XX  
XX Luebbert H;  
PI

XX  
DR WPI: 2001-212797/22.  
DR P-PSDB; AAB67530.  
XX  
PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
PT a transgenic non-human animal as an animal model for neurodegenerative  
PT diseases.  
XX  
PS Claim 3; Page 34-35; 62pp; English.  
XX  
CC The present sequence encodes a murine parkin2 polypeptide. The  
CC polynucleotide sequence contains a frameshift mutation, leading to a  
CC truncated protein. Mutations or deletions in the parkin2 gene cause  
CC Parkinson's disease in humans. The human parkin2 gene is located in gene  
CC region 6q25.2-27. Parkinson2 polypeptides and polynucleotides are useful for  
CC analysing neurodegenerative diseases. They are also useful for testing  
CC the efficacy of the treatment of a neurodegenerative disease such as  
CC Parkinson's disease, Alzheimer's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, multi-system atrophy, Wilson's disease,  
CC Pick's disease, Prion disease, and secondary causes including Parkinson's  
CC syndromes like toxins, drugs, brain tumours, head trauma, stroke,  
CC vascular irregularities or metabolic irregularities, associated with a  
CC less active or non-active parkin protein  
XX

SQ Sequence 3092 BP; 818 A; 778 C; 761 G; 735 T; 0 U; 0 Other;

Query Match 72.2%; Score 340.2; DB 4; Length 3092;  
Best Local Similarity 84.0%; Pred. No. 7e-107;  
Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 11 CAGGTAGATCAATCTACAAAGCTTTATGTATGCAAGAGCCCTGTCAAGAGTGC 70  
DB 375 CAGTTAAACCACTCAACAGCTTTTATCTACAGCAAGAGCCCTGTCAAGAGTGC 434  
QY 71 AGCGGAAAACCTCAGGTCACGTGACGACCTGCGAGGCAAGCTCAGCTTGACCC 130  
DB 435 AGCTTGGAAAGCTCCGAGTTCAGTGTGCGACCTGCAAAACCAAGCAACCTCAGCTTGAGCC 494  
QY 131 AGGTCATCTTGTCTGGAGTGTGATGTTTAAATTCACAAACCGATGAGTGTGAATGCCAAT 190  
DB 495 AGGCGCATCTTGTCTGGACGATGTTTAAATTCACAAACCGATGAGTGTGAATGCCAAT 190  
QY 191 CCCCACTGCTGCTGGAGTGTGACAAATTTTCTTAAATGTGAGACACCCCACT 250  
DB 555 CTCAGACTGCTGGAACCAAGCTGAATTTTCTTAAATGTGAGACACCCCACT 614  
QY 251 CTGACAAAGAAACATCAGTACGTTTGCACCTGATGCAACAAATAGTGGAAATCACTT 310  
DB 615 CAGACAAAGAACGTCGTGATGTTGAACCTGATCACAAGCAAGGCGCAGCATCCCTT 674  
QY 311 GCATTACGTGACAGACGTCAGAGCCCGTCTGTTTCCAGTGCACACTCCGCGCAG 370  
DB 675 GCATTACGTGACAGACGTCAGAGCCCGTCTGTTTCCAGTGCACACTCCGCGCAG 370  
QY 371 TGATTGCTTGAAGCTTTTCCATTATCTGTGTGACAAAGCTCAATGATGGCAGTTTG 430  
DB 735 TGATCTGTTTGAAGCTTTTCCATTATCTGTGTGACAAAGCTCAATGATGGCAGTTTG 430  
QY 431 TTCACGACCCCTCAATGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467  
DB 795 TCCACGATGCTCAATGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831

RESULT 9  
AAFS5256  
ID AAF55256 standard; cDNA, 3253 BP.  
AC AAF55256;  
XX  
DT 29-MAY-2001 (first entry)  
XX  
DE Nucleotide sequence of murine parkin2 with a frameshift mutation.  
XX

Db	776	CAGACAGAGACACGCGCGGTAGCTTTGAACCTGATATCAGCAGCAAGAGCGCAGATCCCTT	835
Qy	311	GCATTACGTGCACAGACGCTCAGAGAGCCCCGCTCTGAGTTTCCAGTGCACCACTCCCGCAGC	370
Db	836	GCATAGCGTGACAGAGATGTACAGAGAGCCCTGCTCGGTCTTCCAGTGTAAACACCCGTCAAG	895
Qy	371	TGATTTCGTTAGACCTGTTTCCACTTATACTGTGTGACAAAGACTCAATGATCGGCACTTTC	430
Db	896	TGATCTGTTTGGACACTGTTTCCACTTGTATTTGTGTACAAAGACTCAACGATCGGCACTTTC	955
Qy	431	TTACAGACCCCTCAACTTGAGTACTCCCTGCGCTTGTGT	467
Db	956	TCCACGATGCTCAACTTGAGTACTCCCTGCGCTGTGT	992
RESULT 10			
AAF55254	ID	AAF55254 standard; cDNA, 3253 BP.	
XX	AC	AAF55254;	
XX	DT	29-MAY-2001 (first entry)	
XX	DE	Nucleotide sequence of murine parkin2 with a frameshift mutation.	
XX	KW	Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;	
KW	Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;		
KW	Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;		
KW	brain tumour; head trauma; stroke; vascular irregularity;		
KW	metabolic irregularity; 88.		
OS	Mus sp.		
XX	FH	Key	Location/Qualifiers
FT	CDS	129..242	
FT		/*tag= a	
FT		/product= "truncated parkin2"	
XX	PN	EP1081225-A1.	
XX	PD	07-MAR-2001.	
XX	PF	30-AUG-1999; 99EP-00116766.	
XX	PR	30-AUG-1999; 99EP-00116766.	
XX	PA	(BIOF-) BIOFRONTIERA PHARM GMBH.	
P1	Luebbert H;		
DR	WPI; 2001-212797/22.		
XX	P-PSDB; AAB67527.		
XX	PT	New polynucleotides encoding mouse parkin2 protein, useful for producing	
XX	PT	a transgenic non-human animal as an animal model for neurodegenerative	
XX	PS	diseases.	
XX	PS	Claim 3; Page 30-31; 62pp; English.	
CC	XX	The present sequence encodes a murine parkin2 polypeptide. The	
CC	XX	polynucleotide sequence contains mutations, causing Gln38 to be changed	
CC	XX	to a stop codon, leading to a truncated protein. Mutations or deletions	
CC	XX	in the parkin gene cause Parkinson's disease in humans. The human	
CC	XX	parkin gene is located in gene region 6q25.2-27. Parkin2 polypeptides	
CC	XX	and polynucleotides are useful for analysing neurodegenerative diseases.	
CC	XX	They are also useful for testing the efficacy of the treatment of a	
CC	XX	neurodegenerative disease such as Parkinson's disease, Alzheimer's	
CC	XX	disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-	
CC	XX	system atrophy, Wilson's disease, Pick's disease, prion disease, and	
CC	XX	secondary causes inducing Parkinson's syndromes like toxins, drugs, brain	
CC	XX	tumours, head trauma, stroke, vascular irregularities or metabolic	
CC	XX	irregularities, associated with a less active or non-active parkin	
CC	XX	protein	

XX Sequence 3253 BP; 851 A; 818 C; 808 G; 776 T; 0 U; 0 Other;  
SQ  
Query Match 72.2%; Score 340.2; DB 4; Length 3253;  
Best Local Similarity 84.0%; Pred. No. 7.2e-107;  
Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 11 CAGGTAGATCAATCTACACAGCTTTTATGTAATGCAAGAGCCCTGTCAAGAGTGC 70  
DB 536 CAGTTAAACCACTACACAGCTTTTATCTACTACTGAAAGGCCCTGCAACAGGTCC 595  
QY 71 AGCCGGGAAAATCTAGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 130  
DB 596 AGCTGGAAAGCTCCGAGTTCAGTGTGACCTGCAAAACAGAACCTTCACTTGGGCC 655  
QY 131 AGGGTCCATCTTGTGGGATGATGTTTAAATTCACAAACCGGATGATGATGATGATG 190  
DB 656 AGGGCCATCTTGTGGGATGATGTTTAAATTCACAAACCGGATGATGATGATGATG 715  
QY 191 CCCCACTGCTGCTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 250  
DB 716 CTCAGACTGCTGCTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 775  
QY 251 CTGACAAAGAAATCATCAGTACCTTTCAGTACCTGATGCAACAAATGTCGAACTCACTT 310  
DB 776 CAGACAAAGAACCTCGGTACCTTTCAGTACCTGATGCAACAAATGTCGAACTCACTT 835  
QY 311 GCATTAGGTGACAGAGTACAGAGCCCGTCTGTTTTCAGTACAGTACAGTACAGTACAGT 370  
DB 836 GCATTAGGTGACAGAGTACAGAGCCCGTCTGTTTTCAGTACAGTACAGTACAGTACAGT 895  
QY 371 TGATTTGTTAGTCTGTTTTCACCTTATCTGTCAGACAGTCAATGATGAGGAGTTTG 430  
DB 896 TGATCTGTTTGAAGTGTTCACCTTATCTGTCAGACAGTCAATGATGAGGAGTTTG 955  
QY 431 TTCACGACCTCAACTGCTGCTACTTCCCTGCTTGTGT 467  
DB 956 TTCACGATGCTCAACTGCTGCTACTTCCCTGCTTGTGT 992  
RESULT 11  
AAFS5255  
ID AAF55255 standard; cDNA; 3254 BP.  
XX  
AC AAF55255;  
XX  
DT 29-MAY-2001 (first entry)  
XX  
DE Nucleotide sequence of murine parkin2 with a frameshift mutation.  
XX  
KW Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
KW Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
KW brain tumour; head trauma; stroke; vascular irregularity;  
KW metabolic irregularity; ss.  
XX  
OS Mus sp.  
XX  
FH Key location/Qualifiers  
FT CDS 129..290  
FT /\*tag= a  
FT /product= "truncated parkin2"  
XX  
PD EP1081225-A1.  
XX  
XX 07-MAR-2001.  
XX  
XX 30-AUG-1999; 99EP-00116766.  
XX  
XX 30-AUG-1999; 99EP-00116766.  
XX  
XX (BIOF-) BIOFRONTIERA PHARM GMBH.  
XX

PI Luebert H;  
XX  
XX WPI: 2001-212797/22.  
DR P-PSDB; AAB67528.  
XX  
PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
PT a transgenic non-human animal as an animal model for neurodegenerative  
PT diseases.  
XX  
PS Claim 3; Page 31-32; 62pp; English.  
XX  
XX The present sequence encodes a murine parkin2 polypeptide. The  
XX polynucleotide sequence contains mutations, causing Asn54 to be changed  
XX to a stop codon, leading to a truncated protein. Mutations or deletions  
XX in the parkin2 gene cause Parkinson's disease in humans. The human  
XX parkin2 gene is located in gene region 6q25.2-27. Parkin2 polypeptides  
XX and polynucleotides are useful for analysing neurodegenerative diseases.  
XX They are also useful for testing the efficacy of the treatment of a  
XX neurodegenerative disease such as Parkinson's disease, Alzheimer's  
XX disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-  
XX system atrophy, Wilson's disease, Pick's disease, Prion disease, and  
XX secondary causes inducing Parkinson's syndromes like toxins, drugs, brain  
XX tumours, head trauma, stroke; vascular irregularities or metabolic  
XX irregularities, associated with a less active or non-active parkin  
XX protein  
SQ Sequence 3254 BP; 851 A; 818 C; 809 G; 776 T; 0 U; 0 Other;  
Query Match 72.2%; Score 340.2; DB 4; Length 3254;  
Best Local Similarity 84.0%; Pred. No. 7.2e-107;  
Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 11 CAGGTAGATCAATCTACACAGCTTTTATGTAATGCAAGAGCCCTGTCAAGAGTGC 70  
DB 537 CAGTTAAACCACTACACAGCTTTTATCTACTACTGAAAGGCCCTGCAACAGTCC 596  
QY 71 AGCCGGGAAAATCTAGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 130  
DB 597 AGCTGGAAAGCTCCGAGTTCAGTGTGACCTGCAAAACAGAACCTTCACTTGGCCC 656  
QY 131 AGGGTCCATCTTGTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 190  
DB 657 AGGGCCATCTTGTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 716  
QY 191 CCCCACTGCTGCTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 250  
DB 717 CTCAGACTGCTGCTGGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 776  
QY 251 CTGACAAAGAAATCATCAGTACCTTTCAGTACCTGATGCAACAAATGTCGAACTCACTT 310  
DB 777 CAGACAAAGAACCTCGGTACCTTTCAGTACCTGATGCAACAAATGTCGAACTCACTT 836  
QY 311 GCATTAGGTGACAGAGTACAGAGCCCGTCTGTTTTCAGTACAGTACAGTACAGTACAGT 370  
DB 837 GCATTAGGTGACAGAGTACAGAGCCCGTCTGTTTTCAGTACAGTACAGTACAGTACAGT 896  
QY 371 TGATTTGTTAGTCTGTTTTCACCTTATCTGTCAGACAGTCAATGATGAGGAGTTTG 430  
DB 897 TGATCTGTTTGAAGTGTTCACCTTATCTGTCAGACAGTCAATGATGAGGAGTTTG 956  
QY 431 TTCACGACCTCAACTGCTGCTACTTCCCTGCTTGTGT 467  
DB 957 TTCACGATGCTCAACTGCTGCTACTTCCCTGCTTGTGT 993  
RESULT 12  
AAFS5259  
ID AAF55259 standard; cDNA; 3255 BP.  
XX  
XX AAF55259;  
XX  
DT 29-MAY-2001 (first entry)  
XX



DE Nucleotide sequence of a mutated murine parkin2 polypeptide.  
 XX  
 KM Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KM Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KM Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KM brain tumour; head trauma; stroke; vascular irregularity;  
 KM metabolic irregularity; ss.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 129..1523  
 FT /\*tag= a  
 FT /product= "truncated parkin2"  
 XX  
 XX EPI081225-A1.  
 PD 07-MAR-2001.  
 XX  
 PF 30-AUG-1999; 99EP-00116766.  
 XX  
 PR 30-AUG-1999; 99EP-00116766.  
 XX  
 PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX  
 PI Luebbert H;  
 XX  
 XX WPI; 2001-212797/22.  
 DR P-PSDB; AAB67532.  
 XX  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PT  
 XX  
 XX Claim 3; Page 36-38; 62pp; English.  
 XX  
 XX The present sequence encodes a murine parkin2 polypeptide. The sequence  
 CC contains the mutation Thr415Asn. Mutations or deletions in the parkin2  
 CC gene cause Parkinson's disease in humans. The human parkin2 gene is  
 CC located in gene region 6q25.2-27. Parkin2 polypeptides and  
 CC polynucleotides are useful for analysing neurodegenerative diseases. They  
 CC are also useful for testing the efficacy of the treatment of a  
 CC neurodegenerative disease such as Parkinson's disease, Alzheimer's  
 CC disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-  
 CC system atrophy, Wilson's disease, Pick's disease, Prion disease, and  
 CC secondary causes inducing Parkinson's syndromes like toxins, drugs, brain  
 CC tumours, head trauma, stroke, vascular irregularities or metabolic  
 CC irregularities, associated with a less active or non-active parkin  
 CC protein.  
 CC  
 XX  
 XX Sequence 3255 BP; 853 A; 817 C; 809 G; 776 T; 0 U; 0 Other;  
 SQ  
 Query Match 72.2%; Score 340.2; DB 4; Length 3255;  
 Best Local Similarity 84.0%; Pred. No. 7.2e-107;  
 Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 11 CAGGTAGTCAATCTACACAGCTTTTATGTGTATGGAAAGCCCTGTAAAGAGTGC 70  
 DB 538 CAGTTAAACCACTACACAGCTTTTATGTGTATGGAAAGCCCTGTACCAAGAGTGC 597  
 QY 71 AGCCGGGAAACTAGAGGTACAGTGCAGCAGCTGAGGCAACGCTGACCTTGAACC 130  
 DB 598 AGCTGGAAAGCTCGAGATTGAGTGTGACCTGCAACCAAGCAACCTTGCCTTGGCCC 657  
 QY 131 AGGGTCCATCTTGTGGGATGATGTTTAAATTCACAAACCGGATGAGTGAATGCCAAT 190  
 DB 658 AGGGCCCATCTTGTGGGAGCATGTCTTAATTCACAAACCGGATGAGTGAATGCCAAT 717  
 QY 191 CCCCACTGCTCCCTGGGAGTGTGAGCAAAATTTTCTTAAATGTGAGCAACACCCCACT 250  
 DB 718 CTCGAGCTGCTCCGGAACCAAGAGTGAATTTTCTTAAATGTGAGCAACACCCCACT 777  
 QY 251 CTGACAGAAACATCAGTACTTGTGACCTGATCGCAACAAATAGTGGAAACATCACTT 310

DB 778 CAGACAGACACGTCGAGTCTTGAACCGATCACAGCAAGGCGGAGCATCCCTT 837  
 QY 311 GCATTACGTGACAGAGTACAGAGCCCGCCGCTGTTTCCATGCAATCCGCCACG 370  
 DB 838 GCATAGGCTGACAGAGTACAGAGCCCGCTGCTGCTTCCAGTGTAAACCAAGTCCG 897  
 QY 371 TGATTGCTTGAAGTCTTTCACCTTATATCTGTGACAAAGTCAATGATCGGAGTTTG 430  
 DB 898 TGATCTGTGAGTCTGTTTCCACTTGTATGTGTGTGACAAAGCTCAACGATGGCAGTTTG 957  
 QY 431 TTCAGACCTCAACTTGGCTACTCCCTGCTGTGT 467  
 DB 958 TCACAGTGTCAACTTGGCTACTCCCTGCTGTGT 994  
 RESULT 13  
 AAF55244  
 ID AAF55244 standard; cDNA; 3255 BP.  
 XX  
 AC AAF55244;  
 XX  
 DT 29-MAY-2001 (first entry)  
 XX  
 XX Nucleotide sequence of a murine parkin2 polypeptide.  
 XX  
 KM Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KM Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KM Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KM brain tumour; head trauma; stroke; vascular irregularity;  
 KM metabolic irregularity; ss.  
 XX  
 XX Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 129..1523  
 FT /\*tag= a  
 FT /product= "parkin2"  
 XX  
 XX EPI081225-A1.  
 PD 07-MAR-2001.  
 XX  
 PF 30-AUG-1999; 99EP-00116766.  
 XX  
 PR 30-AUG-1999; 99EP-00116766.  
 XX  
 PA (BIOF-) BIOFRONTIERA PHARM GMBH.  
 XX  
 PI Luebbert H;  
 XX  
 XX WPI; 2001-212797/22.  
 DR P-PSDB; AAB67517.  
 XX  
 PT New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 PT  
 XX  
 XX Disclosure; Page 15-16; 62pp; English.  
 XX  
 XX The present sequence encodes a murine parkin2 polypeptide. Mutations or  
 CC deletions in the parkin2 gene cause Parkinson's disease in humans. The  
 CC human parkin2 gene is located in gene region 6q25.2-27. Parkin2  
 CC polypeptides and polynucleotides are useful for analysing  
 CC neurodegenerative diseases. They are also useful for testing the efficacy  
 CC of the treatment of a neurodegenerative disease such as Parkinson's  
 CC disease, Alzheimer's disease, Huntington's disease, amyotrophic lateral  
 CC sclerosis, Multi-system atrophy, Wilson's disease, Pick's disease, Prion  
 CC disease, and secondary causes inducing Parkinson's syndromes like toxins,  
 CC drugs, brain tumours, head trauma, stroke, vascular irregularities or  
 CC metabolic irregularities, associated with a less active or non-active  
 CC parkin protein  
 CC

SQ	Sequence	3255 BP; 852 A; 818 C; 809 G; 776 T; 0 U; 0 Other;
	Query Match	72.2%; Score 340.2; DB 4; Length 3255;
	Best Local Similarity	84.0%; Pred. No. 7.2e-107;
	Matches	384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
OY	11	CAGGTGAGTCAATCTACACACAGCTTTTAAGTGATTCGAAGAAGCCCTGTCAAAGATGC 70
Dd	538	CAGTTAAACCACCTTACACAACAGCTTTTTCATCTACTGCAAAAGCCCTGTCCAAGATGC 597
OY	71	AGCCGGGAAAACTCAGGGTACAGTGCAGCACCTGCAGGAGGCAACGCTCACCTGACCC 130
Dd	598	AGCCTGGAAAAGCTCCGAGTTCAAGTGGGACCTCGCAACAGCAACCTCAGCTTGSGCC 657
OY	131	AGGGTTCATCTTGCTGGAGTAGTGTTTAATTCCAAACCGATGAGTGGTGAATCCAAT 190
Dd	658	AGGGGCCCATCTTGCTGGAGCATGTCTTAATTCCAAACCGATGAGTGGTGAATGCCAGT 717
OY	191	CCCCACATCTGCCCCTGGGACTAGTGAGAAATTTTTCTTTAAATGTGAGACACCCCACT 250
Dd	718	CTCCAGACTGCCCCCTGGAACCAAGCTGAATTTTTCTTTAAATGTGAGACACCCCACT 777
OY	251	CTGACAAAGGAACATCATAGTCTTGACCTGTATGCACAAATAATGTGGAAATCATCACTT 310
Dd	778	CAGACAAAGGACACGTCGTGGTAGCTTTGAACCTTATACCAACAGCAGGCGACACATCCCTT 837
OY	311	GCAATACGTGCA CAGACGTCAGAGACCCCGCTCTGGTTTCCAGTGCAACTCCGCGCAG 370
Dd	838	GCATAGCGTGCA CAGATGTCAAGAGCCCGTCTGGTCTTCCAGTGAACCAACCGTCAAG 897
OY	371	TGATTTGCTTAGACTGTTTCCACTTAATCTGTGTGACAAAGACTCAATGATGGCAGTTTG 430
Dd	898	TGATCTGTTTGGACGTTTCCACTTGATTTGTGTACAAAGACTCAACGATGGCGCAGTTTG 957
OY	431	TTCAAGACCCCTCAACTGTGGCTACTCCCGCCCTTGCT 467
Dd	958	TCCACGATGCTTCAACTTGGCTACTCCCGCCGCTGHT 994
RESULT 14		
AAPF55260	ID	AAPF55260 standard; cDNA; 3255 BP.
AC	AAPF55260;	
XX		
DT	29-MAY-2001	(first entry)
XX		
DE	Nucleotide sequence of a mutated murine parkin2 polypeptide.	
XX		
KW	Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;	
KV	Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;	
KW	Multisystem atrophy; Wilson's disease; Pick's disease; Prion disease;	
KW	brain tumour; head trauma; stroke; vascular irregularity;	
KW	metabolic irregularity; ss.	
XX		
OS	Mus sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	129..1484
FT		/*tag= a
FT		/product= "truncated parkin2"
XX		
PN	EP1081225-A1.	
XX		
PD	07-MAR-2001.	
XX		
FP	30-AUG-1999;	99EP-00116766.
XX		
PR	30-AUG-1999;	99EP-00116766.
XX		
PA	(BIOF-) BIOFRONTIERA PHARM GMBH.	
XX		
PI	Luebbert H;	

XX	WPI, 2001-212797/12.
DR	P-PSDB; AAB67533.
XX	
XX	New polynucleotides encoding mouse parkin2 protein, useful for producing
PT	a transgenic non-human animal as an animal model for neurodegenerative
PT	diseases.
XX	
XX	
PS	Claim 3; Page 38-39; 62pp; English.
XX	
CC	The present sequence encodes a murine parkin2 polypeptide. The sequence
CC	contains the mutation Trp453stop. Mutations or deletions in the parkin2
CC	gene cause Parkinson's disease in humans. The human parkin2 gene is
CC	located in gene region 6q25.2-27. Parkin2 polypeptides and
CC	polynucleotides are useful for analysing neurodegenerative diseases. They
CC	are also useful for testing the efficacy of the treatment of a
CC	neurodegenerative disease such as Parkinson's disease, Alzheimer's
CC	disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-
CC	system atrophy, Wilson's disease, Pick's disease, Prion disease, and
CC	secondary causes inducing Parkinson's syndromes like toxins, drugs, brain
CC	tumours, head trauma, stroke, vascular irregularities or metabolic
CC	irregularities, associated with a less active or non-active parkin
CC	protein
XX	
SQ	Sequence 3255 BP; 853 A; 818 C; 808 G; 776 T; 0 U; 0 Other;
XX	
Query Match	72.2%; Score 340.2; DB 4; Length 3255;
Best Local Similarity	84.0%; Pred. No. 7.2e-107;
Matches 384; Conservative	0; Mismatches 73; Indels 0; Gaps 0
QY	11 CAGGTAGTCAATCTTACCAACAGCTTTATGTGTATGTGCMAAGGCCCTGTCAAAGATGC 70
DB	538 CAGTTAAACCCACCTACCAACAGCTTTTCATCTACTGCMAAGGCCCTGTCCAAAGTCC 597
QY	71 AGCCGGGAAAACCTCAGGGGTACAGTCACACACTCGAGGCGAAGCTCAGCTTGACCC 130
DB	598 AGCTCGAAAGCTCCGAGTTCAGTGTGCACCTGCMAAACCAACCAACCTCAGCTTGGCCC 657
QY	131 AGGCTCATCTTCTGCTGGAGTATGTTTTAATTCMAACCGATGAGTGGATGCCAAT 190
DB	658 AGGGGCCATCTTGTCTGGGACGATGCTTAATTCMAACCGATGAGTGGATGCCAGT 717
QY	191 CCCCACTGCCCCCTGGGACTAGTGCAGAAATTTTCTTTAATGTGGAGACACCCCACT 250
DB	718 CTCCAGACTGCCCTGGAAACCAAGAGTGAAATTTTCTTTAATGTGGAGACACCAACT 777
QY	251 CTGACAAAGAAACATCACTACTTGTGCACCTGATGCGAACAAATAGTGGAACTACTT 310
DB	778 CAGCAAAAGAACCGTGGGTACTTTGAACCTGATCACCAGCAACAGGGGCGACATCTCTT 837
QY	311 GCATTACGTGCACAGACGTGCAGAGACCCCGTCCCTGTTTTCAGTGCACCACTCCGCGACG 370
DB	838 GCATATGCGTGCACAGATGTACAGAGCCCTGTGCTGTGCTTTCAGTGTATCAACCGCTACG 897
QY	371 TGAATTTGCTTAGACTGTTTCCACTTATATCTGTGTGACAAAGCTCAATGATGGCAGTTTG 430
DB	898 TGAATCTGTTTGAAGCTGTTTCCACTTGTATGTGTGACAAAGCTCAACGATGGCAGTTTG 957
QY	431 TTCAAGACCTTCAACTTGGCTACTCTCCCTGCTGTGT 467
DB	958 TTCACGATGTCAACTTGGCTACTCTCCCTGCGTGTGT 994
XX	
RESULT 15	
AAFS5258	
ID	AAFS5258 standard; cDNA; 3255 BP.
XX	
XX	AAFS5258;
AC	
DT	29-MAY-2001 (first entry)
XX	
XX	Nucleotide sequence of a mutated murine parkin2 polypeptide.
XX	

KM Parkinson's disease; 6q25.2-27; neurodegenerative disease;  
 KM Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
 KM Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;  
 KM brain tumour; head trauma; stroke; vascular irregularity;  
 KM metabolic irregularity; ss.  
 XX Mus sp.  
 OS  
 PH Key Location/Qualifiers  
 FT CDS 129..1523  
 PT /\*Lag= a  
 PT /product= "truncated parkin2"  
 XX  
 PN BP1081225-A1.  
 PD 07-MAR-2001.  
 XX  
 XX 30-AUG-1999; 99EP-00116766.  
 PF  
 XX 30-AUG-1999; 99EP-00116766.  
 PR  
 XX (BIOF-) BIOFRONTERA PHARM GMBH.  
 PA  
 PI Luebbert H;  
 XX  
 XX WPI: 2001-212797/22.  
 DR P-PSDB; AAB67531.  
 XX  
 XX New polynucleotides encoding mouse parkin2 protein, useful for producing  
 PT a transgenic non-human animal as an animal model for neurodegenerative  
 PT diseases.  
 XX  
 PS Claim 3; Page 35-36; 62pp; English.  
 XX  
 CC The present sequence encodes a murine parkin2 polypeptide. The sequence  
 CC contains the mutation Lys161Asn. Mutations or deletions in the parkin2  
 CC gene cause Parkinson's disease in humans. The human parkin2 gene is  
 CC located in gene region 6q25.2-27. Parkin2 polypeptides and  
 CC polynucleotides are useful for analyzing neurodegenerative diseases. They  
 CC are also useful for testing the efficacy of the treatment of a  
 CC neurodegenerative disease such as Parkinson's disease, Alzheimer's  
 CC disease, Huntington's disease, amyotrophic lateral sclerosis, Multi-  
 CC system atrophy, Wilson's disease, Pick's disease, Prion disease, and  
 CC secondary causes inducing Parkinson's syndromes like toxins, drugs, brain  
 CC tumours, head trauma, stroke, vascular irregularities or metabolic  
 CC irregularities, associated with a less active or non-active parkin  
 CC protein  
 CC  
 XX  
 XX Sequence 3255 BP; 852 A; 818 C; 808 G; 777 T; 0 U; 0 Other;  
 SQ  
 Query Match 72.2%; Score 340.2; DB 4; Length 3255;  
 Best Local Similarity 84.0%; Pred. No. 7.2e-107;  
 Matches 384; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 11 CAGGTATCAATGATACAGCTTTATGTATTTGCAAGGCCCTGTCAAGAGTGC 70  
 DB 538 CAGTTAAACCACTACACAGCTTTTATCTACTGCAAGGCCCTGTCCACAAGTTC 597  
 QY 71 AGCCGGAAAACTCAGGGTACAGTGACGACCTGACGAGGCAACGCTCACTTGACCC 130  
 DB 598 AGCCTGGAATCTCCGAGTTCAGTGTGGCACTGCAAAAGCAACCTCACTTGGGCC 657  
 QY 131 AGGGTCCATCTTGTGGAGTATGTTTAAATCCAAACCGGATGAGTGTGAATGCCAAT 190  
 DB 658 AGGGCCCATCTTGTCTGGAGCATGTCTTAATTCCAAACCGGATGAGTGTGAATGCCAAT 717  
 QY 191 CCCACACTGCCCTGGGAGTATGACAGAAATTTTCTTAAATGTGAGACACACCCCACT 250  
 DB 718 CTCAGACTGCCCTGGAACAGACCTGAATTTTCTTAAATGTGAGACACACCAACT 777  
 QY 251 CTGACAAGAAACATCAGTACCTTGGACCTGATCGCAACAAATAGTCGAACATCACTT 310  
 DB 778 CAGCAAGAGACAGCTCGTAGCTTTGAACCTGATCACGACACAGGCGACGATCCCTT 837

QY 311 GCATTAGTGACAGACGTGAGAGCCCGTCTGTGTTTCCAGTGCNACTCCGCCACG 370  
 DB 838 GCATTAGTGACAGACGTGAGAGCCCGTCTGTGTTTCCAGTGCNACTCCGCCACG 897  
 QY 371 TGATTTGCTTAGACTGTTTCCACTTATATCTGTGTGACAGACTCAATGATGGCAGTTTG 430  
 DB 898 TGATCTGTTGAGACTGTTTCCACTTGTATGTGTACAGACTCAACGATCGGAGTTTG 957  
 QY 431 TTCACGACCTCAACTTGGCTACTCCCTGCTTGTGT 467  
 DB 958 TCACGATGCTCAACTTGGCTACTCCCTGCTGTTGT 994

Search completed: February 25, 2005, 19:01:27  
 Job time : 268.776 secs

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2005, 17:58:29 ; Search time 81.1144 Seconds  
(without alignments)  
9501.226 Million cell updates/sec

Title: US-10-622-817-10

Perfect score: 471

Sequence: 1 ggaagaccagcagtagatc.....actccctccttggtgtag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405668

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/lna/PTCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/lna/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466.4	99.0	2960	4	US-09-601-844B-1
2	466.4	99.0	2960	4	US-09-949-016-645
3	288.4	61.2	2876	4	US-09-601-844B-1
4	288.4	61.2	2876	4	US-09-949-016-982
5	251.8	53.5	2513	4	US-09-949-016-983
6	37.8	8.0	399	4	US-09-621-976-8976
7	34.8	7.4	40	4	US-09-601-844B-65
8	34.8	7.4	361	3	US-09-643-597-293
9	34.8	7.4	361	4	US-09-480-884A-293
10	34.8	7.4	361	4	US-09-542-615A-293
11	34.8	7.4	361	4	US-09-606-421B-293
12	34.8	7.4	405	4	US-09-630-940B-293
13	34.8	7.4	405	4	US-09-513-999C-11324
14	34.8	7.4	40	4	US-09-601-844B-64
15	32.8	7.0	178883	4	US-09-949-016-12733
16	32.8	7.0	178884	4	US-09-949-016-13039
17	32.6	6.9	35262	4	US-09-949-016-12797
18	32.6	6.9	35263	4	US-09-949-016-16399
19	32.2	6.8	152132	4	US-09-949-016-13845
20	32.2	6.8	152145	4	US-09-949-016-12371
21	32.2	6.8	38	4	US-09-601-844B-66
22	32.2	6.8	347	4	US-09-621-976-11665
23	32.2	6.8	819	4	US-09-489-039A-3239
24	31.8	6.8	37	4	US-09-601-844B-68
25	31.8	6.8	342	4	US-09-134-000C-2019
26	31.8	6.8	1160	3	US-09-737-698B-12
27	31.8	6.8	1160	4	US-09-737-626A-12

28	31.8	6.8	1695	3	US-09-737-698B-28	Sequence 28, Appl
29	31.8	6.8	1695	4	US-09-737-626A-28	Sequence 28, Appl
30	31.8	6.8	1830121	4	US-09-557-884-1	Sequence 1, Appl
31	31.8	6.8	1830121	4	US-09-643-990A-1	Sequence 1, Appl
32	31.6	6.7	601	4	US-09-949-016-16816	Sequence 36816, A
33	31.6	6.7	601	4	US-09-949-016-18953	Sequence 58953, A
34	31.6	6.7	601	4	US-09-949-016-184559	Sequence 184559, A
35	31.6	6.7	119801	4	US-09-949-016-13453	Sequence 13453, A
36	31.6	6.7	190078	4	US-09-949-016-12707	Sequence 12707, A
37	31.6	6.7	190078	4	US-09-949-016-17026	Sequence 17026, A
38	31.6	6.7	312470	4	US-09-949-016-14043	Sequence 14043, A
39	31.6	6.7	336024	4	US-09-949-016-12373	Sequence 12373, A
40	31.4	6.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
41	31.2	6.6	27727	4	US-09-949-016-15737	Sequence 15737, A
42	31.2	6.6	29067	4	US-09-949-016-12535	Sequence 12535, A
43	31.2	6.6	29093	4	US-09-949-016-17476	Sequence 17476, A
44	31.2	6.6	74940	4	US-09-949-016-17512	Sequence 17512, A
45	30.8	6.5	7194	4	US-09-601-326-76	Sequence 76, Appl

## ALIGNMENTS

RESULT 1  
US-09-601-844B-1  
; Sequence 1, Application US/09601844B  
; Patent No. 6716621  
; GENERAL INFORMATION:  
; APPLICANT: Shimizu, Yoshiaki  
; TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease  
; FILE REFERENCE: 0652.2110000  
; CURRENT APPLICATION NUMBER: US/09/601,844B  
; PRIOR APPLICATION NUMBER: 2000-08-09  
; PRIOR FILING DATE: 1999-02-09  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 2960  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (102)..(1496)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (102)..(108)  
; OTHER INFORMATION: Exon 1  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (109)..(272)  
; OTHER INFORMATION: Exon 2  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (273)..(513)  
; OTHER INFORMATION: Exon 3  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (514)..(635)  
; OTHER INFORMATION: Exon 4  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (636)..(719)  
; OTHER INFORMATION: Exon 5  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (720)..(835)  
; OTHER INFORMATION: Exon 6  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (836)..(972)

```
/ OTHER INFORMATION: Exon 7
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (973)..(1034)
/ OTHER INFORMATION: Exon 8
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1035)..(1184)
/ OTHER INFORMATION: Exon 9
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1185)..(1268)
/ OTHER INFORMATION: Exon 10
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1269)..(1386)
/ OTHER INFORMATION: Exon 11
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1387)..(2960)
/ OTHER INFORMATION: Exon 12
US-09-601-844B-1
```

```
Query Match 99.0%; Score 466.4; DB 4; Length 2960;
Best Local Similarity 99.8%; Pred. No. 1.9e-154;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GGAAGTCCGACGAGTATGATCAATCTTAAACAGCTTTTATGATGGAAGGCCCCCTGT 60
DB 504 GGAAGTCCGACGAGTATGATCAATCTTAAACAGCTTTTATGATGGAAGGCCCCCTGT 563
QY 61 CAAAGAGTGACGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGCAAGCAAGCTC 120
DB 564 CAAAGAGTGACGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGCAAGCAAGCTC 623
QY 121 ACCTTGACCCAGGGTCCATCTTGTGAGATGATGTTTAAATCCAAACCGGATGATGCT 180
DB 624 ACCTTGACCCAGGGTCCATCTTGTGAGATGATGTTTAAATCCAAACCGGATGATGCT 683
QY 181 GAATGCCAATCCCACTGCGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 240
DB 684 GAATGCCAATCCCACTGCGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 743
QY 241 CACCCCACTCTGACAGAGAAACATCAGTACCTTGCACCTGATCGCAACAAATAGTCGG 300
DB 744 CACCCCACTCTGACAGAGAAACATCAGTACCTTGCACCTGATCGCAACAAATAGTCGG 803
QY 301 AACATCACTTGACATTAAGTGCAGACGTCAGAGCCCGCTGCTGTTTCCAGTGCAC 360
DB 804 AACATCACTTGACATTAAGTGCAGACGTCAGAGCCCGCTGCTGTTTCCAGTGCAC 863
QY 361 TCCCGCAGCTGATTTGCTTACAGCTGTTCCATTATCTGTGACAGACTCAATGAT 420
DB 864 TCCCGCAGCTGATTTGCTTACAGCTGTTCCATTATCTGTGACAGACTCAATGAT 923
QY 421 CGGAGTTTGTTCACAGACCTCAACTGCTACTCCCTGCTGTGTG 468
DB 924 CGGAGTTTGTTCACAGACCTCAACTGCTACTCCCTGCTGTGTG 971
```

```
RESULT 2
US-09-949-016-645
/ Sequence 645; Application US/09949016
/ Patent No. 681239
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
```

```
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 645
/ LENGTH: 2960
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-645
```

```
Query Match 99.0%; Score 466.4; DB 4; Length 2960;
Best Local Similarity 99.8%; Pred. No. 1.9e-154;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GGAAGTCCGACGAGTATGATCAATCTTAAACAGCTTTTATGATGGAAGGCCCCCTGT 60
DB 504 GGAAGTCCGACGAGTATGATCAATCTTAAACAGCTTTTATGATGGAAGGCCCCCTGT 563
QY 61 CAAAGAGTGACGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGCAAGCAAGCTC 120
DB 564 CAAAGAGTGACGCCGGGAAAACCTCAGGGTACAGTGCAGACCTGCAGGCAAGCAAGCTC 623
QY 121 ACCTTGACCCAGGGTCCATCTTGTGAGATGATGTTTAAATCCAAACCGGATGATGCT 180
DB 624 ACCTTGACCCAGGGTCCATCTTGTGAGATGATGTTTAAATCCAAACCGGATGATGCT 683
QY 181 GAATGCCAATCCCACTGCGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 240
DB 684 GAATGCCAATCCCACTGCGCTGGGACCTAGTGCAGAAATTTTCTTAAATGTGAGCA 743
QY 241 CACCCCACTCTGACAGAGAAACATCAGTACCTTGCACCTGATCGCAACAAATAGTCGG 300
DB 744 CACCCCACTCTGACAGAGAAACATCAGTACCTTGCACCTGATCGCAACAAATAGTCGG 803
QY 301 AACATCACTTGACATTAAGTGCAGACGTCAGAGCCCGCTGCTGTTTCCAGTGCAC 360
DB 804 AACATCACTTGACATTAAGTGCAGACGTCAGAGCCCGCTGCTGTTTCCAGTGCAC 863
QY 361 TCCCGCAGCTGATTTGCTTACAGCTGTTCCATTATCTGTGACAGACTCAATGAT 420
DB 864 TCCCGCAGCTGATTTGCTTACAGCTGTTCCATTATCTGTGACAGACTCAATGAT 923
QY 421 CGGAGTTTGTTCACAGACCTCAACTGCTACTCCCTGCTGTGTG 468
DB 924 CGGAGTTTGTTCACAGACCTCAACTGCTACTCCCTGCTGTGTG 971
```

```
RESULT 3
US-09-601-844B-3
/ Sequence 3; Application US/09601844B
/ Patent No. 6716621
/ GENERAL INFORMATION:
/ APPLICANT: Shimizu, No. 6716621uyoshi
/ APPLICANT: Mizuno, Yoshikuni
/ TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease
/ FILE REFERENCE: 0652.211000
/ CURRENT APPLICATION NUMBER: US/09/601,844B
/ PRIOR FILING DATE: 2000-08-09
/ PRIOR APPLICATION NUMBER: PCT/JP99/00545
/ PRIOR FILING DATE: 1999-02-09
/ NUMBER OF SEQ ID NOS: 70
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 2876
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (102)..(1412)
/ OTHER INFORMATION:
/ FEATURE:
```

```
NAME/KEY: misc feature
LOCATION: (102)..(108)
OTHER INFORMATION: Exon 1
FEATURE:
NAME/KEY: misc feature
LOCATION: (109)..(272)
OTHER INFORMATION: Exon 2
FEATURE:
NAME/KEY: misc feature
LOCATION: (273)..(513)
OTHER INFORMATION: Exon 3
FEATURE:
NAME/KEY: misc feature
LOCATION: (514)..(635)
OTHER INFORMATION: Exon 4
FEATURE:
NAME/KEY: misc feature
LOCATION: (636)..(751)
OTHER INFORMATION: Exon 6
FEATURE:
NAME/KEY: misc feature
LOCATION: (752)..(888)
OTHER INFORMATION: Exon 7
FEATURE:
NAME/KEY: misc feature
LOCATION: (889)..(950)
OTHER INFORMATION: Exon 8
FEATURE:
NAME/KEY: misc feature
LOCATION: (951)..(1100)
OTHER INFORMATION: Exon 9
FEATURE:
NAME/KEY: misc feature
LOCATION: (1101)..(1184)
OTHER INFORMATION: Exon 10
FEATURE:
NAME/KEY: misc feature
LOCATION: (1185)..(1302)
OTHER INFORMATION: Exon 11
FEATURE:
NAME/KEY: misc feature
LOCATION: (1303)..(2876)
OTHER INFORMATION: Exon 11
US-09-601-844B-3
```

Query Match 61.2%; Score 288.4; DB 4; Length 2876;  
Best local Similarity 81.8%; Pred. No. 2.7e-91;  
Matches 383; Conservative 0; Mismatches 1; Indels 84; Gaps 1;

```
QY 1 GGAAGTCCAGAGGTAAGTCAATCTACAACAGCTTTTATGTGATTTGCAAAAGGCCCTGT 60
DB 504 GGAAGTCCAGAGGTAAGTCAATCTACAACAGCTTTTATGTGATTTGCAAAAGGCCCTGT 563
QY 61 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGCACTCGCAAGCAACGCTC 120
DB 564 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGCACTCGCAAGCAACGCTC 623
QY 121 ACCTTGACCCAGGGTCCATCTTGCTGGAGTATGTTTAATTCAAACCGGATGATGCT 180
DB 624 ACCTTGACCCAG-----GATTTTCTTTAAATGTGAGCA 635
QY 181 GAATGCCAATCCCAACACTGCGCTGGACTAGTGAGAAATTTTCTTTAAATGTGAGCA 240
DB 636 -----GATTTTCTTTAAATGTGAGCA 659
QY 241 CACCCCACTCTGCACAAAGAAACATCAGTAGCTTTGCACCTGATCGCAACAAATAGTCGG 300
DB 660 CACCCCACTCTGCACAAAGAAACATCAGTAGCTTTGCACCTGATCGCAACAAATAGTCGG 719
QY 301 AACATCACTTGCAATTAAGTGCACAGACGTCAGAGACCCCGTCTGCTTTTCCAGTGCAC 360
DB 720 AACATCACTTGCAATTAAGTGCACAGACGTCAGAGACCCCGTCTGCTTTTCCAGTGCAC 779
```

```
QY 361 TCCCGCAGCGATTTGCTTAGACTGTTTCCACTTATACGTGTGACAAAGCTCAATGAT 420
DB 780 TCCCGCAGCGATTTGCTTAGACTGTTTCCACTTATACGTGTGACAAAGCTCAATGAT 839
QY 421 CGGCAATTTGTTACAGACCCCTCAACTTGGCTACTCCCTGCTTGTGTG 468
DB 840 CGGCAATTTGTTACAGACCCCTCAACTTGGCTACTCCCTGCTTGTGTG 887
```

RESULT 4  
US-09-949-016-982  
Sequence 982, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949, 016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: PASCSEQ for Windows Version 4.0  
SEQ ID NO 982  
LENGTH: 2876  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-982

Query Match 61.2%; Score 288.4; DB 4; Length 2876;  
Best local Similarity 81.8%; Pred. No. 2.7e-91;  
Matches 383; Conservative 0; Mismatches 1; Indels 84; Gaps 1;

```
QY 1 GGAAGTCCAGAGGTAAGTCAATCTACAACAGCTTTTATGTGATTTGCAAAAGGCCCTGT 60
DB 504 GGAAGTCCAGAGGTAAGTCAATCTACAACAGCTTTTATGTGATTTGCAAAAGGCCCTGT 563
QY 61 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGCACTCGCAAGCAACGCTC 120
DB 564 CAAAGAGTGCAGCCGGGAAAACCTCAGGGTACAGTGCAGCACTCGCAAGCAACGCTC 623
QY 121 ACCTTGACCCAGGGTCCATCTTGCTGGAGTATGTTTAATTCAAACCGGATGATGCT 180
DB 624 ACCTTGACCCAG-----GATTTTCTTTAAATGTGAGCA 635
QY 181 GAATGCCAATCCCAACACTGCGCTGGACTAGTGAGAAATTTTCTTTAAATGTGAGCA 240
DB 636 -----GATTTTCTTTAAATGTGAGCA 659
QY 241 CACCCCACTCTGCACAAAGAAACATCAGTAGCTTTGCACCTGATCGCAACAAATAGTCGG 300
DB 660 CACCCCACTCTGCACAAAGAAACATCAGTAGCTTTGCACCTGATCGCAACAAATAGTCGG 719
QY 301 AACATCACTTGCAATTAAGTGCACAGACGTCAGAGACCCCGTCTGCTTTTCCAGTGCAC 360
DB 720 AACATCACTTGCAATTAAGTGCACAGACGTCAGAGACCCCGTCTGCTTTTCCAGTGCAC 779
QY 361 TCCCGCAGCGATTTGCTTAGACTGTTTCCACTTATACGTGTGACAAAGCTCAATGAT 420
DB 780 TCCCGCAGCGATTTGCTTAGACTGTTTCCACTTATACGTGTGACAAAGCTCAATGAT 839
QY 421 CGGCAATTTGTTACAGACCCCTCAACTTGGCTACTCCCTGCTTGTGTG 468
DB 840 CGGCAATTTGTTACAGACCCCTCAACTTGGCTACTCCCTGCTTGTGTG 887
```

RESULT 5  
US-09-949-016-983

```

: Sequence 983, Application US/09949016
: Patent No. 6812339
:
: GENERAL INFORMATION:
:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CLO01307
:
: CURRENT APPLICATION NUMBER: US/09/949,016
:
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
:
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
:
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
:
: PRIOR FILING DATE: 2000-03-08
:
: NUMBER OF SEQ ID NOS: 207012
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 983
:
: LENGTH: 2513
:
: TYPE: DNA
:
: ORGANISM: Human
:
: US-09-949-016-983

```

Query Match	53.5%	Score 251.8	DB 4	length 2513	
Best Local Similarity	97.3%	Pred. No. 2,4e-78			
Matches	256	Conservative	0	Mismatches	7
				Indels	0
				Gaps	0
QY	206	GGACTAGTGCAGAAATTTTCTTTAAATGTGGAGCACACCCCACTCTGACAGAAACAT	265		
Db	262	GGACTGTGCAGAAATTTTCTTTAAATGTGGAGCACACCCCACTCTGACAGAAACAC	321		
QY	266	CAGTAGCTTTGCACCTGATTCGCAACAATATGTCGGAATCATCTGCAATTGCTGCACAG	325		
Db	322	CAGTAGCTTTGCACCTGATTCGCAACAATATGTCGGAATCATCTGCAATTGCTGCACAG	381		
QY	326	ACGTTCAGAGCCCCGCTCTGTGTTTCCAGTGCACACTCCCGCAGGTGATTTGCTTAGACT	385		
Db	382	ACGTTCAGAGCCCCGCTCTGTGTTTCCAGTGCACACTCCCGCAGGTGATTTGCTTAGACT	441		
QY	386	GTTCCTCAATTATCTGTGTGACAAGACATCAATGATCGCAGTTTGTTACGACCCCTCAAC	445		
Db	442	GTTCCTCAATTATCTGTGTGACAAGACATCAATGATCGCAGTTTGTTACGACCCCTCAAC	501		
QY	446	TTGGCTACTCCCTGACCTTTGTGTG	468		
Db	502	TTGGCTACTCCCTGACCTTTGTGTGTG	524		

```

RESULT 621
US-09-621-976-8976
; Sequence 6976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins
; FILE REFERENCE: GENSET, 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-8976

```

Query Match	8.0%;	Score 37.8;	DB 4;	Length 399;
Best Local Similarity	12.0%;	Pred.No. 0.0069;		
Matches	36;	Conservative 134;	Mismatches 131;	Indels 0;
				Gaps 0;
166	AACCGAGATGGTGGATGCCAATCCCAACACTGCGCCCTGGACATGTCAGAAATTTTTC			225

Db 20 RAYGMYTTRSSKSMWAKSMKKRRRRRRAMWWWKSMCKKSKSWRSWGMYTTRKMG 79

Qy 226 TTTTAAATGTGAGACACCCCACTCTGACAAGAAACATCAGTAGCTTGACCTGATC 285

Db 80 RGAASWAGYMSWMTYTRMRYYRRKKACTKWAAGMGAGAMAAVAKMYMMARRT 139

Qy 286 GCAACAATATGTGGAAACATCATCTTGCAATTACGTGCACACAGCTCAGAGACCCCGTC 345

Db 140 AMKYMMMSKSRMRRRRMYAMTYMMARRTMGMRASCYRGAYMASAGYMMWYMM 199

Qy 346 GTTTCACAGTCACTCCCGCACCTGATTTGCTTAAGCTGTTTCCACTATATCTGNGT 405

Db 200 RRMKMYTSAGMSMRKTRRCASYSWSSYCMNGAKMMYMTKTSNISTWYSYCTKRRS 259

Qy 406 ACAAGACTCAATGATGCGCATGTTGTTCACGACCTCACTGTGGCTATCCTCCGCTGTT 465

Db 260 CCGMSMSCYWKTYRYSWYCASCYYSYKTKPRASCMMCCMKMRMMAMMYMMQVYCKKT 319

Qy 466 G 466

Db 320 S 320

```

RESULT 7
US-09-601-844B-65
: Sequence 65, Application US/09601844B
: Parent No. 6716621
: GENERAL INFORMATION:
: APPLICANT: Shimizu, Yoshiaki
: APPLICANT: Mizuno, Yoshikuni
: TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease
: FILE REFERENCE: 0652.211000
: CURRENT APPLICATION NUMBER: US/09/601.844B
: CURRENT FILING DATE: 2000-08-09
: PRIOR APPLICATION NUMBER: PCT/JP99/00545
: PRIOR FILING DATE: 1999-02-09
: NUMBER OF SEQ ID NOS: 70
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 65
: LENGTH: 40
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-601-844B-65

```

```

Query Match      7.4%; Score 34.8; DB 4; Length 40;
Best Local Similarity 94.7%; Pred. No. 0.02;
Matches 36; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      128 CCGAGGTCATCTTGCTGGGATCATGTTTTAATTCCA 165
          |||
Db       3 CCAAGGTCATCTTGTGGGATCATGTTTTAATTCCA 40

```

RESULT 8  
US-09-643-597-293  
Sequence 293, Application US/09643557  
Patent No. 6426072  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hoeken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.455C11



;; CURRENT APPLICATION NUMBER: US/09/643,597  
;; CURRENT FILING DATE: 2000-08-21  
;; NUMBER OF SEQ ID NOS: 369  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 293  
;; LENGTH: 361  
;; TYPE: DNA  
;; ORGANISM: Homo sapien  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(361)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-643-597-293

Query Match 7.4%; Score 34.8; DB 3; Length 361;  
Best Local Similarity 59.4%; Pred. No. 0.076;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 203 CTGGAGCTAGTGCAGAAATTTTCTTTAAATGTGAGACACCCCACTCTGACAAGAA 262  
DB 134 CTTGAACACGCTTCAATTTTCTTCCAAATGCTGATGCCACCTTGAGTAAGAA 193  
QY 263 CATCAGTACCTTTGACCTGATGCAACAATAATGTC 298  
DB 194 CANAAGTATTTTAAACATGACAGCTAANMAACATTC 229

RESULT 9  
US-09-480-884A-293  
; Sequence 293, Application US/09480884A  
; Patent No. 6482597  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.455C6  
; CURRENT APPLICATION NUMBER: US/09/480,884A  
; CURRENT FILING DATE: 2001-08-27  
; NUMBER OF SEQ ID NOS: 330  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 293  
; LENGTH: 361  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(361)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-480-884A-293

Query Match 7.4%; Score 34.8; DB 4; Length 361;  
Best Local Similarity 59.4%; Pred. No. 0.076;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 203 CTGGAGCTAGTGCAGAAATTTTCTTTAAATGTGAGACACCCCACTCTGACAAGAA 262  
DB 134 CTTGAACACGCTTCAATTTTCTTCCAAATGCTGATGCCACCTTGAGTAAGAA 193  
QY 263 CATCAGTACCTTTGACCTGATGCAACAATAATGTC 298  
DB 194 CANAAGTATTTTAAACATGACAGCTAANMAACATTC 229

RESULT 10  
US-09-542-615A-293  
; Sequence 293, Application US/09542615A  
; Patent No. 6518256  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong

;; APPLICANT: Fan, Liqun  
;; APPLICANT: Kalos, Michael D.  
;; APPLICANT: Bangur, Chaitanya S.  
;; APPLICANT: Hosken, Nancy A.  
;; APPLICANT: Fanger, Gary R.  
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
;; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.455C8  
;; CURRENT APPLICATION NUMBER: US/09/542,615A  
;; CURRENT FILING DATE: 2000-04-14  
;; NUMBER OF SEQ ID NOS: 350  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 293  
;; LENGTH: 361  
;; TYPE: DNA  
;; ORGANISM: Homo sapien  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(361)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-542-615A-293

Query Match 7.4%; Score 34.8; DB 4; Length 361;  
Best Local Similarity 59.4%; Pred. No. 0.076;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 203 CTGGAGCTAGTGCAGAAATTTTCTTTAAATGTGAGACACCCCACTCTGACAAGAA 262  
DB 134 CTTGAACACGCTTCAATTTTCTTCCAAATGCTGATGCCACCTTGAGTAAGAA 193  
QY 263 CATCAGTACCTTTGACCTGATGCAACAATAATGTC 298  
DB 194 CANAAGTATTTTAAACATGACAGCTAANMAACATTC 229

RESULT 11  
US-09-606-421B-293  
; Sequence 293, Application US/09606421B  
; Patent No. 6531315  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Hosken, Nancy  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skelky, Yasir A.W.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.455C9  
; CURRENT APPLICATION NUMBER: US/09/606,421B  
; CURRENT FILING DATE: 2000-06-28  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 293  
; LENGTH: 361  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(361)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-606-421B-293

Query Match 7.4%; Score 34.8; DB 4; Length 361;  
Best Local Similarity 59.4%; Pred. No. 0.076;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 203 CTGGAGCTAGTGCAGAAATTTTCTTTAAATGTGAGACACCCCACTCTGACAAGAA 262  
DB 134 CTTGAACACGCTTCAATTTTCTTCCAAATGCTGATGCCACCTTGAGTAAGAA 193

QY 263 CATCAGTAGCTTTGCACCTGATCCCAACAAATAGTC 298  
 Db 194 CANNAGTATTTTAAACATGACACGCTAANAACATTC 229

RESULT 12  
US-09-630-940B-293  
; Sequence 293, Application US/09630940B  
; Patent No. 6737514

```

GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chantanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C10
CURRENT APPLICATION NUMBER: US/09/630,940B
CURRENT FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 367
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 293
LENGTH: 361
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(361)
OTHER INFORMATION: n = A,T,C or G
US-09-630-940B-293

```

	Query Match	Similarity	Score	DB	Length	Best Local	Mismatches	Conservative	Indels	Gaps
	7.4%	59.4%	34.8	DB 4	361		0	39	0	0
Qy	203	CTGGGACATGAGCAGAAATTTTCTTTAATGTGAGACACCCCACTGACCAAGAA	262							
Db	134	CTTGACACGCTTCAATTTTCTTCCAAATGCTGCATGCCACCTTGAGTAACGAG	193							
Qy	263	CATCAGTAGCTTTGACCTGATTCGACAAATAGTC	298							
Db	194	CANAAGTATTTTAAACATGACAGCTAANAACTTC	229							

```

RESULT 13
US-09-513-999C-11324/C
; Sequence 11324, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J. B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J. Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 11324
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 326
; OTHER INFORMATION: w=a or t
US-09-513-999C-11324

```

```
Query Match      7.4%; Score 34.8; DB 4; Length 405;
Best Local Similarity 59.4%; Pred. No. 0.081;
Matches 57; Conservative 1; Mismatches 38; Indels 0; Gaps 0;
```

Qy	Db
203 CTGGGACCTGATCAGAAATTTTCTTTAAATGTGGAGCAGACCCACCTGACAGAGAAA	262
333 CTTGAACWCGCTCTTCAATTTTCTTCCAAAATGCTGATGCACACTTGGAGTAAAGAA	274
Qy	Db
263 CATCACTAGCTTTCGACCTGATCCGAAACAAATATGTC	298
273 CAGAAATGATTTTAAACATGACAGGTAAAGACATTC	238

RESULT 14  
 US-09-601-844B-64  
 : Sequence 64. Application US/09601844B  
 : Patent No. 6716621  
 : GENERAL INFORMATION:  
 : APPLICANT: Shimizu, Yoshiaki  
 : APPLICANT: Mizuno, Yoshikuni  
 : TITLE OF INVENTION: Isolated DNA or Gene Responsible for Parkinson's Disease  
 : FILE REFERENCE: 0652.2110000  
 : CURRENT APPLICATION NUMBER: US/09/601,844B  
 : CURRENT FILING DATE: 2000-08-09  
 : PRIOR APPLICATION NUMBER: PCT/JP99/00545  
 : PRIOR FILING DATE: 1999-02-09  
 : NUMBER OF SEQ ID NOS: 70  
 : SOFTWARE: PatentIn version 3.1  
 : SEQ ID NO 64  
 : LENGTH: 40  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : US-09-601-844B-64

```

Query Match 7.2% Score 34; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 131 AGGATCATCTTGTCGGATGATGATTTTAATTC 164
      |||||
Db 7 AGGATCATCTTGTCGGATGATGATTTTAATTC 40

```

```

RESULT 15
US-09-949-016-12733/C
; Sequence 12733, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12733
; LENGTH: 178863
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12733

```

Query Match 7.0%; Score 32.8; DB 4; Length 178883;  
Best Local Similarity 54.0%; Pred. No. 17;  
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy	33	CTTTATGTTGTAATGCAAGGCCCTGTCAAGAGTGCAGCCGGAAACTCAGGGTACA	92
Db	109301	CTTGTCGGTGTGATGCAATGTGCTGGGTTTGCAAAATGTGACTCACAGCCAGCGGCTTACA	109242
Qy	93	GTGCAGCACTGCAGGCGAAGGCTCACTTGAACCCAGGGTCACTTGTGGGATGA	152
Db	109241	AACCTAGTGTGCGAGGAGAAACAGTCATCCAGAAATCAGGGCACAGTTGCCGTGGTGC	109182
Qy	153	TGTT	156
Db	109181	TGCT	109178

Search completed: February 26, 2005, 05:20:30  
Job time : 88.114 secs

***This Page Blank (uspio)***



```
Db 121 ACCTGACCCAGGGTCCATCTTGCGATGATGATTTTAATTCAAACGGATAGTGT 180
Qy 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
Db 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
Qy 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 300
Db 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 300
Qy 301 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 360
Db 301 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 360
Qy 361 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Db 361 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Qy 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 471
Db 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 471
```

## RESULT 2

```
US-10-622-817-10
; Sequence 10, Application US/10622817
; Publication No. US20040214763A1
; GENERAL INFORMATION:
; APPLICANT: CORTEI, Olga
; APPLICANT: HAMPE, Cornelia
; APPLICANT: BRICE, Alexis
; APPLICANT: ROONEY, Laurent
; APPLICANT: FOURNIER, Alain
; TITLE OF INVENTION: METHOD FOR DETERMINING THE ABILITY OF A COMPOUND TO MODIFY THE
; FILE REFERENCE: FRAV2002-0020 US NP
; CURRENT APPLICATION NUMBER: US/10/622,817
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: GB 0229934.5
; PRIOR FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/396,929
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(471)
US-10-622-817-10
```

Query Match 100.0%; Score 471; DB 18; Length 471;  
Best Local Similarity 100.0%; Pred. No. 9.5e-157;  
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 GGAATGTCAGAGTAGATCAATCTACAAGCTTTTATGTGATGCAAAAGCCCTGT 60
Db 1 GGAATGTCAGAGTAGATCAATCTACAAGCTTTTATGTGATGCAAAAGCCCTGT 60
Qy 61 CAAAGAGTCAGAGCGGGAACCTCAGGGTACAGTGCAGACCTGACAGCAAGCGTC 120
Db 61 CAAAGAGTCAGAGCGGGAACCTCAGGGTACAGTGCAGACCTGACAGCAAGCGTC 120
Qy 121 ACCTGACCCAGGGTCCATCTTGCTGAGTATGTTTAAATTCAAACCGGATAGTGT 180
Db 121 ACCTGACCCAGGGTCCATCTTGCTGAGTATGTTTAAATTCAAACCGGATAGTGT 180
Qy 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
Db 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
```

```
Db 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
Qy 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 300
Db 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 300
Qy 301 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 360
Db 301 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 360
Qy 361 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Db 361 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Qy 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 471
Db 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 471
```

## RESULT 3

```
US-10-839-688-12
; Sequence 12, Application US/10839688
; Publication No. US20050014173A1
; GENERAL INFORMATION:
; APPLICANT: Farret, Matthew J.
; TITLE OF INVENTION: PARKINSON'S DISEASE MARKERS
; FILE REFERENCE: 07039-448001
; CURRENT APPLICATION NUMBER: US/10/839,688
; PRIOR FILING DATE: 2004-05-05
; PRIOR APPLICATION NUMBER: US 60/468,832
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2955
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-839-688-12
```

Query Match 99.0%; Score 466.4; DB 19; Length 2955;  
Best Local Similarity 99.8%; Pred. No. 1.1e-154;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 1 GGAATGTCAGAGTAGATCAATCTACAAGCTTTTATGTGATGCAAAAGCCCTGT 60
Db 499 GGAATGTCAGAGTAGATCAATCTACAAGCTTTTATGTGATGCAAAAGCCCTGT 558
Qy 61 CAAAGAGTCAGAGCGGGAACCTCAGGGTACAGTGCAGACCTGACAGCAAGCGTC 120
Db 559 CAAAGAGTCAGAGCGGGAACCTCAGGGTACAGTGCAGACCTGACAGCAAGCGTC 618
Qy 121 ACCTGACCCAGGGTCCATCTTGCTGAGTATGTTTAAATTCAAACCGGATAGTGT 180
Db 619 ACCTGACCCAGGGTCCATCTTGCTGAGTATGTTTAAATTCAAACCGGATAGTGT 678
Qy 181 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 240
Db 679 GAATGCCAATCCCCACACTGCTCCCTGGAGCTAGTGCAGAAATTTTAAATGTGAGCA 738
Qy 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 300
Db 739 CACCCACCTCTGACAAAGAAACATCAGTACTTTGCACTGATGCAACAAATAGTCGG 798
Qy 301 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 360
Db 799 AACATCACTTGCAATTAACGTGACAGAGCTGAGAGCCCGCTGCTTCCAGTCAAC 858
Qy 361 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Db 859 TCCCGCACGTGATTTGCTTAGACTGTTTCCACTTAATCTGTGACAAAGTCAATGAT 420
Qy 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 468
Db 421 CGGAGTTTGTTCAGACGCCCTCAACTTGCTACTCCCTGCTGTGTGTAG 468
```



PRIOR FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 60/396,929  
PRIOR FILING DATE: 2002-07-16  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 7  
LENGTH: 2960  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (102)..(1499)  
US-10-622-817-7

Query Match  
Best Local Similarity 99.0%; Score 466.4; DB 18; Length 2960;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 60  
DB 504 GGAAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 60  
QY 61 CAAAGAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 563  
DB 564 CAAAGAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 563  
QY 121 ACCTTGACCCAGGATCCATCTTGTGGGATGATGTTTAAATTCGCAACCGGATGATGTT 180  
DB 624 ACCTTGACCCAGGATCCATCTTGTGGGATGATGTTTAAATTCGCAACCGGATGATGTT 180  
QY 181 GAATGCCAATCCCACTGCTGCTGGGAGTATGTCAGAAATTTTCTTTAAATGTGAGCA 240  
DB 684 GAATGCCAATCCCACTGCTGCTGGGAGTATGTCAGAAATTTTCTTTAAATGTGAGCA 240  
QY 241 CACCCCACTCTGCAAGAAACATCAGTAGCTTTGCACTGATGCAACCAATAGTGG 743  
DB 744 CACCCCACTCTGCAAGAAACATCAGTAGCTTTGCACTGATGCAACCAATAGTGG 743  
QY 301 AACATCACTTGCATTAAGTGCACAGCTCAGGAGCCCGCTGCTTTTCCAGTGCAC 360  
DB 804 AACATCACTTGCATTAAGTGCACAGCTCAGGAGCCCGCTGCTTTTCCAGTGCAC 360  
QY 361 TCCCGCAGCTGATTTGCTTGAAGCTTTTCACTTATCTGTGTGCAAGAGCTCAATGAT 420  
DB 864 TCCCGCAGCTGATTTGCTTGAAGCTTTTCACTTATCTGTGTGCAAGAGCTCAATGAT 420  
QY 421 CGGAGTTTGTTCACGACCTTCACTTGTGCTTCTGCTGCTTGTGTG 468  
DB 924 CGGAGTTTGTTCACGACCTTCACTTGTGCTTCTGCTGCTTGTGTG 971

RESULT 7  
US-10-776-604-1  
Sequence 1, Application US/10776604  
Publication No. US2005000385A1  
GENERAL INFORMATION:  
APPLICANT: Shimizu, Nobuyoshi  
APPLICANT: Mizuno, Yoshikuni  
TITLE OF INVENTION: DNAs or Genes Participating in Parkinson's Disease  
FILE REFERENCE: 0652.2110001  
CURRENT FILING DATE: 2004-02-12  
PRIOR APPLICATION NUMBER: US/10/776,604  
PRIOR FILING DATE: 2000-08-09  
PRIOR APPLICATION NUMBER: PCT/JP99/00545  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: JP 10/27531  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 2960  
TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (102)..(1496)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (102)..(108)  
OTHER INFORMATION: Exon 1  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (109)..(1272)  
OTHER INFORMATION: Exon 2  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (273)..(513)  
OTHER INFORMATION: Exon 3  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (514)..(635)  
OTHER INFORMATION: Exon 4  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (636)..(719)  
OTHER INFORMATION: Exon 5  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (720)..(835)  
OTHER INFORMATION: Exon 6  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (836)..(972)  
OTHER INFORMATION: Exon 7  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (973)..(1034)  
OTHER INFORMATION: Exon 8  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1035)..(1184)  
OTHER INFORMATION: Exon 9  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1185)..(1268)  
OTHER INFORMATION: Exon 10  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1269)..(1386)  
OTHER INFORMATION: Exon 11  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1387)..(2960)  
OTHER INFORMATION: Exon 12  
US-10-776-604-1

Query Match  
Best Local Similarity 99.0%; Score 466.4; DB 18; Length 2960;  
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 60  
DB 504 GGAAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 563  
QY 61 CAAAGAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 563  
DB 564 CAAAGAGTCCAGCAGGTGATCAATCTACAACCTTTTATGTGTATTTGCAAGGCCCTGT 563  
QY 121 ACCTTGACCCAGGATCCATCTTGTGGGATGATGTTTAAATTCGCAACCGGATGATGTT 180  
DB 624 ACCTTGACCCAGGATCCATCTTGTGGGATGATGTTTAAATTCGCAACCGGATGATGTT 180  
QY 181 GAATGCCAATCCCACTGCTGCTGGGAGTATGTCAGAAATTTTCTTTAAATGTGAGCA 240



Dh 684 GAATGCCAATCCCACTGCGCTGGAGTAGTGAGAAATTTCTTAAATGTGAGCA 743  
Qy 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 300  
Db 744 CACCCACCTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 803  
Qy 301 AACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 360  
Db 804 AACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 863  
Qy 361 TCCCGCAGCGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGAT 420  
Db 864 TCCCGCAGCGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGAT 923  
Qy 421 CGGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 468  
Db 924 CGGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 971

## RESULT 8

US-10-839-688-11  
; Sequence 11, Application US/10839688  
; Publication No. US20050014173A1  
; GENERAL INFORMATION:  
; APPLICANT: Farret, Matthew J.  
; TITLE OF INVENTION: PARKINSON'S DISEASE MARKERS  
; FILE REFERENCE: 07039-448001  
; CURRENT APPLICATION NUMBER: US/10/839, 688  
; CURRENT FILING DATE: 2004-05-05  
; PRIOR APPLICATION NUMBER: US 60/468,832  
; PRIOR FILING DATE: 2003-05-08  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: PastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 2960  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-839-688-11

Query Match 99.0%; Score 466.4; DB 19; Length 2960;  
Best Local Similarity 99.8%; Pred. No. 1.1e-154;

Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGAAGTCCAGAGGTAGATCAATCTACACAGCTTTTATGTGTAAGAGCCCTG 60  
Db 504 GGAAGTCCAGAGGTAGATCAATCTACACAGCTTTTATGTGTAAGAGCCCTG 563  
Qy 61 CAAAGATGACAGCCGGGAAAACCTCAGGGTACAGTGACAGCTGACAGGCAACGCTC 120  
Db 564 CAAAGATGACAGCCGGGAAAACCTCAGGGTACAGTGACAGCTGACAGGCAACGCTC 623  
Qy 121 ACCTTGACCCAGGGTCCATCTTGTGCGGATGATGTTTAAATCCAAACCGGATGATG 180  
Db 624 ACCTTGACCCAGGGTCCATCTTGTGCGGATGATGTTTAAATCCAAACCGGATGATG 683  
Qy 181 GAATGCCAATCCCACTGCGCTGGAGCTAGTGAGAAATTTCTTAAATGTGAGCA 240  
Db 684 GAATGCCAATCCCACTGCGCTGGAGCTAGTGAGAAATTTCTTAAATGTGAGCA 743  
Qy 241 CACCCACCTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 300  
Db 744 CACCCACCTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 803  
Qy 301 AACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 360  
Db 804 AACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 863  
Qy 361 TCCCGCAGCGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGAT 420  
Db 864 TCCCGCAGCGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGAT 923  
Qy 421 CGGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 468

Dh 924 CGGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 971

## RESULT 9

US-10-622-817-13  
; Sequence 13, Application US/10622817  
; Publication No. US20040214763A1  
; GENERAL INFORMATION:  
; APPLICANT: CORTE, Olga  
; APPLICANT: HAMPE, Cornelia  
; APPLICANT: BRICE, Alexis  
; APPLICANT: PRADIER, Laurent  
; APPLICANT: ROONEY, Thomas  
; APPLICANT: FOURNIER, Alain  
; TITLE OF INVENTION: METHOD FOR DETERMINING THE ABILITY OF A COMPOUND TO MODIFY THE  
; FILE REFERENCE: PRAV2002-0020 US NP  
; CURRENT APPLICATION NUMBER: US/10/622,817  
; PRIOR APPLICATION NUMBER: GB 0229934.5  
; PRIOR FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: US 60/396,929  
; PRIOR FILING DATE: 2002-07-18  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 1470  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1470)  
US-10-622-817-13

Query Match 73.7%; Score 347; DB 18; Length 1470;  
Best Local Similarity 83.9%; Pred. No. 2.7e-112;  
Matches 392; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 2 GAATCCAGAGGTAGATCAATCTACACAGCTTTTATGTGTAAGAGCCCTG 61  
Db 476 GAATCCAGAGGTAGATCAATCTACACAGCTTTTATGTGTAAGAGCCCTG 535  
Qy 62 AAAGATGACAGCCGGGAAAACCTCAGGGTACAGTGACAGCTGACAGGCAACGCTCA 121  
Db 536 AAAGATGACAGCCGGGAAAACCTCAGGGTACAGTGACAGCTGACAGGCAACGCTCA 595  
Qy 122 CTTGACCCAGGGTCCATCTTGTGCGGATGATGTTTAAATCCAAACCGGATGATG 181  
Db 596 CTTGACCCAGGGTCCATCTTGTGCGGATGATGTTTAAATCCAAACCGGATGATG 655  
Qy 182 AATGCCAATCCCACTGCGCTGGAGCTAGTGAGAAATTTCTTAAATGTGAGCAC 241  
Db 656 AATGCCAATCCCACTGCGCTGGAGCTAGTGAGAAATTTCTTAAATGTGAGCAC 715  
Qy 242 ACCCACTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 301  
Db 716 ACCCACTCTGACAAAGAAACATCAGTACTTTGACCTGATCGCAAAATAGTCG 775  
Qy 302 ACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 361  
Db 776 ACATCACTTGCATTAAGTCAGACAGCTGAGAGCCCGCTGCTGTTTCCAGTGCAAC 835  
Qy 362 CCGGCAAGGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGATC 421  
Db 836 CCGGCAAGGATTTGCTTGAAGTCTTTTCCATTAACTGTGACAAAGCTCAATGATC 895  
Qy 422 GCGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 468  
Db 896 GCGAGTTTGTTCAGACCTCACTTGCTGCTACCTCCGCTTGCTG 942

RESULT 10  
US-10-776-604-3

```
Sequence 3, Application US/10776604
Publication No. US2005000385A1
GENERAL INFORMATION:
APPLICANT: Shimizu, Nobuyoshi
APPLICANT: Mizuno, Yoshikuni
TITLE OF INVENTION: DNAs or Genes Participating in Parkinson's Disease
FILE REFERENCE: 0652.2110001
CURRENT APPLICATION NUMBER: US/10/776,604
CURRENT FILING DATE: 2004-02-12
PRIORITY APPLICATION NUMBER: 09/601,844
PRIORITY FILING DATE: 2000-08-09
PRIORITY APPLICATION NUMBER: PCT/JP99/00545
PRIORITY FILING DATE: 1999-02-09
PRIORITY APPLICATION NUMBER: JP 10/27531
PRIORITY FILING DATE: 1998-02-09
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 2876
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (102)..(1412)
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102)..(108)
OTHER INFORMATION: Exon 1
FEATURE:
NAME/KEY: misc_feature
LOCATION: (109)..(272)
OTHER INFORMATION: Exon 2
FEATURE:
NAME/KEY: misc_feature
LOCATION: (273)..(513)
OTHER INFORMATION: Exon 3
FEATURE:
NAME/KEY: misc_feature
LOCATION: (514)..(635)
OTHER INFORMATION: Exon 4
FEATURE:
NAME/KEY: misc_feature
LOCATION: (636)..(751)
OTHER INFORMATION: Exon 6
FEATURE:
NAME/KEY: misc_feature
LOCATION: (752)..(888)
OTHER INFORMATION: Exon 7
FEATURE:
NAME/KEY: misc_feature
LOCATION: (889)..(950)
OTHER INFORMATION: Exon 8
FEATURE:
NAME/KEY: misc_feature
LOCATION: (951)..(1100)
OTHER INFORMATION: Exon 9
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1101)..(1184)
OTHER INFORMATION: Exon 10
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1185)..(1302)
OTHER INFORMATION: Exon 11
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1303)..(2876)
OTHER INFORMATION: Exon 12
US-10-776-604-3
```

```
Query Match 61.2%; Score 288.4; DB 18; Length 2876;
Best Local Similarity 81.8%; Pred. No. 2.9e-91;
```

```
Matches 383; Conservative 0; Mismatches 1; Indels 84; Gaps 1;
QY 1 GGAACTCCAGCAGGTAGATCAATCTACAACAGCTTTATGTGATGCAAAAGGCCCTGT 60
DB 504 GGAACTCCAGCAGGTAGATCAATCTACAACAGCTTTATGTGATGCAAAAGGCCCTGT 563
QY 61 CAAGAAGTCAGCCCGGGAAAACTAGGGTACAGTGACACCTGACGAGGCAACGCTC 120
DB 564 CAAGAAGTCAGCCCGGGAAAACTAGGGTACAGTGACACCTGACGAGGCAACGCTC 623
QY 121 ACCTTGACCCAGGGTTCATCTGTGCGGATGATGTTTAATTCAAACCGGATGAGTGT 180
DB 624 ACCTTGACCCAGG-----GAATTTTCTTTAATGTGGAGCA 635
QY 181 GAATGCCAATCCCACTGCTGCGGAGTACAGAGATTTTCTTTAATGTGAGCA 240
DB 636 -----GAATTTTCTTTAATGTGGAGCA 659
QY 241 CACCCCACTCTGACAGAGAAACATCAGTCTTTGACCTGATCCGACAAATAGTCG 300
DB 660 CACCCCACTCTGACAGAGAAACATCAGTCTTTGACCTGATCCGACAAATAGTCG 719
QY 301 AACATCACTTGACATTAAGTCGACAGACGTGAGAGCCCGTCTGGTTTCCAGTCAAC 360
DB 720 AACATCACTTGACATTAAGTCGACAGACGTGAGAGCCCGTCTGGTTTCCAGTCAAC 779
QY 361 TCCCGCCAGGTGATTTGCTTAAGCTGTTTCCATTATACGTGTGACAAAGACTCAATGAT 420
DB 780 TCCCGCCAGGTGATTTGCTTAAGCTGTTTCCATTATACGTGTGACAAAGACTCAATGAT 839
QY 421 CGGAGTTTGTTCACGACCTCAACTGGTACTCCCTGCTTGTTG 468
DB 840 CGGAGTTTGTTCACGACCTCAACTGGTACTCCCTGCTTGTTG 887

RESULT 11
US-10-839-688-6
Sequence 6, Application US/10839688
Publication No. US2005001417A1
GENERAL INFORMATION:
APPLICANT: Farrer, Matthew J.
TITLE OF INVENTION: PARKINSON'S DISEASE MARKERS
FILE REFERENCE: 07039-448001
CURRENT APPLICATION NUMBER: US/10/839,688
CURRENT FILING DATE: 2004-05-05
PRIORITY APPLICATION NUMBER: US 60/468,832
PRIORITY FILING DATE: 2003-05-08
NUMBER OF SEQ ID NOS: 81
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 650
TYPE: DNA
ORGANISM: Homo sapiens
US-10-839-688-6

Query Match 29.4%; Score 138.6; DB 19; Length 650;
Best Local Similarity 99.3%; Pred. No. 3.4e-38;
Matches 138; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 330 CAGGAGCCCCGTCCTGTTTTCAGTCACTCCGCGCAGTGAATTTGCTTAGACTGTTT 389
DB 208 CAGGAGCCCCGTCCTGTTTTCAGTCACTCCGCGCAGTGAATTTGCTTAGACTGTTT 267
QY 390 CCACTTATACGTGTGACAAAGACTCAATGATCGGAGTTGTTCACGACCCCTCAACTTG 449
DB 268 CCACTTATACGTGTGACAAAGACTCAATGATCGGAGTTGTTCACGACCCCTCAACTTG 327
QY 450 CTACTCCCTGCTTGTGTG 468
DB 328 CTACTCCCTGCTTGTGTG 346

RESULT 12
```



Db	134	CTTGACACGCTTCAATTTCTTCCAAATGCTGCATGCCAOCCTTGAGTACGAAG	193
Qy	263	CATCAGTAGCTTTGACCTGATGCGACAAATAGTC	298
Db	194	CANAAGTATTTTAAACATGACAGCTAANAACAATTC	229

Search completed: February 26, 2005, 13:19:16  
Job time : 286.738 secs